FDRE TECHNICAL AND VOCATIONAL TRAINING INSTITUTE

(FDRE TVTI)

DEPARTMENT OF INFORMATION TECHNOLOGY

THESIS TITLE:

ALIGNING TECHNICAL AND VOCATIONAL EDUCATION AND TRAINING (TVET) WITH LABOR MARKET DEMANDS IN WEST WELLEGA ZONE THROUGH PREDICTIVE DATA ANALYTICS

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

By:

Kibiru Bulti Gudina

ID No: TTMR/138/16

Principal Advisor: Prof. Ravindra Babu (PhD)

Addis Ababa, Ethiopia

June, 2026

FDRE TECHNICAL AND VOCATIONAL TRAINING INSTITUTE
COLLEGE OF GRADUATE STUDIES
DEPARTMENT OF INFORMATION TECHNOLOGY

ALIGNING TECHNICAL AND VOCATIONAL EDUCATION AND TRAINING (TVET) WITH LABOR MARKET DEMANDS IN WEST WELLEGA ZONE THROUGH PREDICTIVE DATA ANALYTICS

By: Kibiru Bulti Gudina
ID No: TTMR/138/16

A thesis submitted to the Department of Information Technology, College of Graduate Studies, FDRE Technical and Vocational Training Institute (FDRE TVTI), in partial fulfillment of the requirements for the degree of Master of Science in Information Technology.

Principal Advisor: Prof. Ravindra Babu (PhD)

Addis Ababa, Ethiopia

June, 2026

DECLARATION

I, the undersigned, declare that this thesis entitled "Aligning Technical and Vocational Education and Training (TVET) with Labor Market Demands in West Wellega Zone through Predictive Data Analytics" is my own original work, prepared under the supervision of my advisor(s). It has not been submitted for the award of any degree, diploma, or any other academic qualification in this or any other institution. All sources of materials used in this thesis have been duly acknowledged.

Declared by:

Name: Kibiru Bulti Gudina

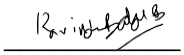
Signature: 

Department: Information Technology, FDRE TVTI

Date: 04/06/2026

This thesis has been submitted for examination with my approval as the candidate's advisor.

Principal Advisor: Prof. Ravindra Babu (PhD)

Signature: 

Date: 04/06/2026

APPROVAL SHEET

This is to certify that the thesis entitled "Aligning Technical and Vocational Education and Training (TVET) with Labor Market Demands in West Wellega Zone through Predictive Data Analytics" submitted by Kibiru Bulti Gudina (ID No. TTMR/138/16), in partial fulfillment of the requirements for the degree of Master of Science in Information Technology, has been examined and approved by the Board of Examiners of the Department of Information Technology, FDRE Technical and Vocational Training Institute.

Examination Role	Name	Signature	Date
Principal Advisor	Prof. Ravindra Babu (PhD)	_____	_____
Co-Advisor (if any)	_____	_____	_____
Internal Examiner	_____	_____	_____
External Examiner	_____	_____	_____
Department Head, IT	_____	_____	_____
Dean, College of Graduate Studies	_____	_____ –	_____

Place: Addis Ababa, Ethiopia

Date of Final Approval: _____

ACKNOWLEDGEMENT

First and above all, I would like to thank God for granting me the strength, wisdom, and determination to conduct this research. My principal advisor, Prof. Ravindra Babu (PhD), deserves great appreciation for his guidance, comments, and support throughout the research. I am grateful to the Department of Information and Communication Technology and the Federal TVET Institute for their support and academic resources during my studies. I am thankful to the faculty members, my friends, and colleagues for their cooperation. I would also like to express my gratitude to the TVET institutions, employers, government officials, and graduates of the West Wellega Zone for their cooperation. I am grateful to them for sharing their opinions and experiences, which helped me, complete my research. My sincere thanks go to my family and friends for their encouragement and support during this exciting and challenging journey. Finally, I would like to extend my sincere thanks to all those who directly and indirectly helped me complete this research.

ABSTRACT

This study explored the role of predictive data analytics in aligning technical and vocational education and training (TVET) programs with labor market demand in West Wellega Zone, Ethiopia. This study was conducted in response to concerns that Technical and Vocational Education and Training (TVET) graduates face unemployment or underemployment due to institutional planning that lacks integrated labor market information. This study analyzed a secondary labor market dataset comprising 2,450 records obtained from the West Wellega Zone Job Creation and Skills Office for the period 2020-2025. During data preprocessing, the dataset was checked for missing values and duplicates (none were found), categorical attributes were encoded as numeric values, and numerical attributes were normalized. Sector, skill, qualification requirement, experience years, graduates, field of study, vacancies, wage level, and year were used as predictor variables to classify the demand level into three classes (Low, Moderate, and High). In the multi-class classification of labor demand, the performances of three supervised machine learning models, Logistic Regression, Decision Tree, and Random Forest, were compared. K-means clustering was applied to group skills and occupations based on similar levels of demand. The evaluation metrics used to assess the models included accuracy, precision, recall, F1-score, and confusion matrices. The Random Forest model outperformed all other models with an accuracy of 92.04%, a precision of 93.07%, a recall of 92.04%, and an F1-score of 92.2%. The Decision Tree yielded an accuracy of 91.84%, and the Logistic Regression model resulted in an accuracy of 89.59%. K-means clustering resulted in identifying four different clusters in the labor market that require different TVET-related inputs. The results show that vacancies, graduate supply, and sector and wage-related demand data can provide relevant evidence for TVET planning, using the right predictive models. The novelty of this study lies in being the first to combine multi-class supervised classification with unsupervised skill clustering for evidence-based TVET planning at the zonal level in rural Ethiopia, a context not addressed by prior national or urban studies.

Keywords: *TVET, labor market demand, predictive data analytics, logistic regression, decision tree, random forest, K-means clustering,* skills mismatch, multi-class classification.

Table of Contents

DECLARATION i
APPROVAL SHEET ii
ACKNOWLEDGEMENT iii
ABSTRACT iv
LIST OF TABLES ix
LIST OF FIGURES x
LIST OF ABBREVIATIONS AND ACRONYMS xi
CHAPTER ONE: INTRODUCTION 1
1.1. Background of the Study 2
1.2. Motivation of the Study 3
1.3. Statement of the Problem 3
1.4. Research Questions 4
1.5. Objectives of the Study 4
1.5.1. General Objective 4
1.5.2. Specific Objectives 4
1.6. Scope of the Study 4
1.7. Significance of the Study 5
1.8. Contribution of the Study 6
1.9. Limitations of the Study 6
1.10. Organization of the Study 7
CHAPTER TWO: REVIEW OF RELATED LITERATURE 8
2.1. Overview of Technical and Vocational Education and Training 8
2.2. Theoretical Literature Review 8
2.3. Digital Transformation and the Future of Work in TVET 12
2.4. Empirical Literature Review 12
2.5. Related Works 17
2.6. Conceptual Framework 18
2.6.1. Diagram of the Conceptual Framework 20
2.7. Research Gap Analysis 20
2.8. Literature Review Summary 22
CHAPTER THREE: RESEARCH METHODOLOGY 23
3.1. Overview 23

3.2. Research Design 23
3.3. Study Area 24
3.4. Data Collection Methods 24
3.4.1. Ethical Consideration 24
3.5. Data Source and Description 24
3.6. Target Class 25
3.7. Data pre-processing 25
3.7.1. Handling Missing Values 26
3.7.2. Encoding Categorical Variables 26
3.7.3. Feature Scaling 26
3.7.4. SMOTE Balancing Technique 27
3.7.5. Feature Importance 27
3.7.6. Feature Selection 28
3.7.7. Train-Test Split 28
3.8. Predictive Model Development 29
3.9. Model Selection and Training 30
3.9.1. Random Forest 30
3.9.2. Logistic Regression 31
3.9.3. Decision Tree 31
3.9.4. K-means Clustering 31
3.10. Cross-Validation 31
3.11. Model Evaluation 32
3.11.1. Confusion Matrix 32
3.11.2. Accuracy 32
3.11.3. Precision 33
3.11.4. Recall (Sensitivity) 33
3.11.5. F1-Score 34
3.12. Development Tools and Technologies 34
3.13. Summary 34
CHAPTER FOUR: RESULTS AND DISCUSSION 35
4.1. Introduction 35
4.2. Exploratory Data Analysis 35
4.2.1. Descriptive Analysis 35

4.2.2. Correlation Analysis 36
4.3. Current Alignment between TVET Training and Labor Market Demands 37
4.3.1. Overview of Existing TVET Programs in West Wellega Zone 38
4.4. Availability and Use of Labor Market Information (LMI) 39
4.5. Historical Labor Market Demand Trends 39
4.6. Implementation Details 40
4.6.1. Hyperparameter Tuning 40
4.7. Predictive Modeling of Labor Demand 41
4.7.1. Introduction to Predictive Modeling 41
4.7.2. Random Forest Model Results and Interpretation 42
4.7.2.1. Error Analysis and Policy Interpretation 44
4.7.3. Logistic Regression Model Results and Interpretation 45
4.7.4. Decision Tree Model Results and Interpretation 47
4.8. Clustering Analysis Using K-Means and Principal Component Analysis 48
4.8.1. Cluster Formation and Validation 49
4.8.2. Elbow Method (WCSS Analysis) 49
4.8.3. Silhouette Score Analysis 50
4.8.4. Cluster Validation Metrics 50
4.8.5. Cross-Validation Results (5-Fold Stratified) 50
4.8.6. Cluster Visualization and Interpretation 51
4.8.7. Ranking the Skills 52
4.8.8. Policy Recommendations by Cluster 53
4.9. Model Comparison Results and Discussion 54
4.9.1. Model Performance 54
4.9.2. Performance Comparison 55
4.10. Results and Discussion 56
4.10.1. Interpretation of Results 57
4.10.2. Implications for TVET Alignment 58
4.10.3. Implications of Predictive Analytics Results 58
CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS 60
5.1. Conclusion 60
5.2. Recommendations 60
5.2.1. Strengthening Labor Market Information Systems (LMIS) 61

5.2.2. Enhancing Industry-TVET Collaboration.....61
5.2.3. Curriculum Reform Based on Labor Market Data 61
5.2.4. Development of a Predictive Analytics-Based TVET Planning Framework 61
5.3. Recommendations for Future Researchers.....62
5.4. Policy Implications.....62
REFERENCES 63
Appendix.....71
Appendix A: Sample Dataset 71
Appendix B: Data Dictionary.....71
Appendix C: Ethical Approval Letter.....72

LIST OF TABLES

Table 1: Summary of Related Work 21
Table 2: Description of Predictor Variables (Input Features) 25
Table 3: Missing Values after Cleaning 26
Table 4: Stratified Split Distribution 29
Table 5: Chronological Data Split by year 29
Table 6: Model Selection based on their complementary strengths 30
Table 7: Summary of Tools and Technologies Used in the Study 34
Table 8: Descriptive Statistics 36
Table 9: Hyperparameter Tuning Results 40
Table 10: Sample dataset used for experiment 41
Table 11: Confusion Matrix of Random Forest 43
Table 12: Definition of Error Types 45
Table 13: Confusion Matrix of Logistic Regression Model 46
Table 14: Confusion Matrix of Decision Tree Model 48
Table 15: 5-Fold Cross-Validation Results (Mean ± Standard Deviation) 51
Table 16: Top 10 Skills Ranking 52
Table 17: Policy Recommendation by Cluster 53
Table 18: Model Comparison 55
Table 19: Complete Data Dictionary for Labor Market Dataset 71

LIST OF FIGURES

Figure 1: Conceptual Framework of TVET Training with Labor Market Demands Alignment 20
Figure 3: Proposed Model Architecture for TVET-Labor Market Alignment 23
Figure 4: Class Distribution Before SMOTE 27
Figure 5: Class Distribution After SMOTE 27
Figure 6: Feature Importance from Random Forest Model 28
Figure 7: Correlation Matrix of Selected Features 37
Figure 8: Confusion Matrix for Random Forest Demand Classification 43
Figure 9: Confusion Matrix of Logistic Regression Model 46
Figure 10: Confusion Matrix of Decision Tree Model 47
Figure 11: Elbow Method for Optimal K 49
Figure 12: Silhouette Score for Different K 50
Figure 13: Skill Clustering based on demand and Wage 51
Figure 14: Top 10 skills by frequency 53

LIST OF ABBREVIATIONS AND ACRONYMS

Abbreviation	Full Form
TVET	Technical and Vocational Education and Training
LMI	Labor Market Information
LMIS	Labor Market Information System
MoLS	Ministry of Labor and Skills
NPAJC	National Plan of Action for Job Creation
EBDM	Evidence-Based Decision-Making
DDDM	Data-Driven Decision-Making
ICT	Information and Communication Technology
SMEs	Small and Medium Enterprises
CEDEFOP	European Centre for the Development of Vocational Training
ILO	International Labor Organization
AI	Artificial Intelligence
ML	Machine Learning
NITA	National Industrial Training Authority
CSV	Comma-Separated Values
EDA	Exploratory Data Analysis
RF	Random Forest
DT	Decision Tree
LR	Logistic Regression
PCA	Principal Component Analysis
SMOTE	Synthetic Minority Over-sampling Technique
WCSS	Within-Cluster Sum of Square
FDRE	Federal Democratic Republic of Ethiopia
TVTI	Technical and Vocational Training Institute
CBET	Competency-Based Education and Training

CHAPTER ONE: INTRODUCTION

Technical and Vocational Education and Training (TVET) is critically important for preparing young people for participation in the labor market by equipping them with skills that are currently in demand. Germany and Singapore are frequently cited as exemplars of effective TVET systems that integrate classroom instruction with structured workplace learning [1]. In Ethiopia, and particularly in the West Wellega Zone, TVET is regarded as a key means of reducing youth unemployment and supporting local economic growth. Despite this potential, a significant challenge remains in the gap between the skills taught in TVET institutions and the evolving needs of employers in the labor market.

This mismatch contributes to underemployment and inefficient allocation of training resources and inefficient resource use in both the education and labor sectors. The proposed project has supported the Ministry of Labor and Skills (MoLS) Technical and Vocational Education and Training (TVET) strategic shift and the National Plan of Action for Jobs Creation (NPAJC) skills development strategies to integrate technology, enhance gender-sensitive training provision, promote green programs, and ensure more competent, demand-driven skills training programs. It aims to improve the quality and relevance of skills development programs through labor market-responsive short-term training to produce a competent workforce and strengthen the supply and demand side linkages to reduce skills mismatch and ensure proficiency in soft, technical, and ICT-based skill sets.

Technical Vocational and Technical Education (TVET) has the potential to influence the labor market and transform the economy, resulting in an improvement in the advancement of careers [2]. Technical and Vocational Education and Training (TVET) is an educational approach aimed at imparting knowledge, skills, and competencies essential for specific occupations to meet the demands of today's dynamic labor market [3]. One of the biggest challenges for education institutes in the design of a labor market-related curriculum is to make sure that the curriculum is up-to-date and uses the same vocabulary as the labor market. This may sound easy, but almost any curriculum is written in academic language and, at the same time, may lack terms that are currently used in the labor market, which may prevent students from appropriately presenting the skills they learn at a TVET institution in the job market. One of the roles of the Department of Manpower in placing workers and expanding job opportunities is to submit job vacancy

information according to company needs [4]. The labor market in the West Wellega Zone is evolving due to several socio-economic and technological shifts. Data analytics is a useful way to deal with this problem. Policymakers and educators can make better decisions by systematically collecting and analyzing information about the job market, such as hiring trends, new industry needs, wage trends, and vacancy data [5]. Predictive analytics tools can provide TVET institutions with labor market intelligence that is almost real-time, which lets them keep their training programs up to date and improve them as the needs of the workforce change.

1.1. Background of the Study

For several decades, Technical and Vocational Education and Training (TVET) has equipped labor forces with skills that enable entry into job markets while contributing to economic development and poverty reduction in developing countries. In a developing country like Ethiopia, Technical and Vocational Education and Training (TVET) plays a crucial role in combating unemployment and poverty. TVET systems impart skills that reflect labor market demand [6]. In Ethiopia, TVET has a vital role in combating youth unemployment and contributing to productivity and industrialization. The government has made tremendous efforts to expand TVET institutions and make them responsive to national demand by launching the National TVET Strategy [7].

The skills offered by TVET programs and those required by the labor market continue to diverge despite these efforts. In Ethiopia's West Wellega Zone, where a variety of industries, including agriculture, informal commerce, small-scale manufacturing, construction, and services, power the economy, this disparity is particularly noticeable. Employers frequently complain that graduates lack industry-specific, problem-solving, and practical abilities, making it difficult for them to find jobs that match their training.

The insufficient utilization of reliable labor market data to guide TVET development is a major factor in this discrepancy. Conventional techniques, such as expert consultations and periodic assessments, are frequently out of date, fragmented, and unable to capture the fluctuations of the labor market in real time. As a result, training programs usually don't take into account the demands of the sector, either now or in the future [8]. This results in wasted resources and high rates of underemployment and unemployment among graduates.

Without data-informed planning, TVET institutions cannot effectively adapt to changing workforce demands, leading to a misallocation of public resources. As a result, certain occupations may be overcrowded with graduates while others remain vacant, affecting the economic potential and career prospects of young individuals [9]. There is a need to explore innovative solutions, such as predictive analytics for data-informed planning and forecasting, to produce relevant data on the labor market that can guide TVET institutions to meet changing workforce demands.

1.2. Motivation of the Study

TVET graduates in Ethiopia face underemployment and unemployment due to a mismatch between their skills and job market needs. This is due to limited labor market information, no linkages to industry, and traditional planning methods that do not respond to change. In regions like the West Wellega Zone, there is no localized research, further hindering training matching. Thus, this study is motivated by the need to use predictive data analytics to better match TVET training to labor market needs, increase employability, and support regional economic growth.

1.3. Statement of the Problem

Technical and Vocational Education and Training (TVET) plays a significant role in human capital development in the West Wellega Zone, Ethiopia. However, irrespective of the expansion of TVET institutions, the issue of training outputs meeting labor market needs has not been addressed properly for three interrelated reasons.

First, current TVET program planning follows a supply-driven approach based on past enrollment trends and occasional employer discussions. Without a forecasting mechanism, TVET institutions risk producing graduates for low-demand fields while missing opportunities to supply emerging high-demand sectors.

Second, labor market data in the zone is fragmented across multiple government offices and not systematically integrated for TVET planning. Data on vacancies, graduate supply, wage trends, and employment outcomes are collected independently but not synthesized into actionable intelligence.

Third, there is no empirical study on the application of predictive data analytics for TVET planning in rural Ethiopia, specifically in West Wellega Zone, Oromia Regional State. Existing studies focus on national and urban areas. This study aims to fill these gaps by

utilizing a labor market data-based predictive data analytics approach for evidence-based TVET planning.

1.4. Research Questions

1. What is the current level of alignment between selected TVET training programs and labor market demand in West Wellega Zone?
2. Which labor market variables are most relevant for supporting data-driven TVET planning?
3. Which predictive model performs best in classifying TVET-labor market alignment or demand level?
4. Which predictors have the greatest influence on labor market demand classification?
5. How can predictive analytics outputs be translated into curriculum planning and LMIS improvement?

1.5. Objectives of the Study

1.5.1. General Objective

To develop and evaluate a predictive data analytics approach for supporting the alignment of TVET training programs with labor market demand in the West Wellega Zone.

1.5.2. Specific Objectives

- To assess the current alignment between selected TVET training programs and labor market demand in the West Wellega Zone.
- To identify and structure relevant labor market variables required for data-driven TVET planning.
- To develop and evaluate predictive models for classifying TVET-labor market alignment or demand level.
- To identify the most influential predictors of labor market demand and interpret their implications for TVET curriculum planning.
- To propose evidence-based recommendations for strengthening LMIS, curriculum reform, and industry-TVET collaboration.

1.6. Scope of the Study

This study explores the relationship between Technical and Vocational Education and Training (TVET) and labor market needs in the West Wellega Zone based on labor market information from 2020 to 2025. This study uses descriptive statistics, Logistic

regression, Decision Tree, K-means clustering, and random forest modeling. The data from the secondary labor market data from the selected region were used. This study is limited by the focus on the West Wellega Zone. It may not fully reflect other regions. The analyses cover a few sectors and data sets, which may not include all occupations. Qualitative factors like issues around the institution, policy implementation, or social and cultural aspects of TVET do not enter into the study.

1.7. Significance of the Study

This study provides a data-driven approach for improved alignment of Technical and Vocational Education and Training (TVET) requirements with the labor market in the West Wellega Zone. It employs predictive data analysis methods such as Logistic Regression, Decision Tree, K-means clustering, and Random Forest to provide insights into current and future skill requirements.

1. **For TVET Institutions**

TVET training can be more effective when using predictive data analytics to analyze student outcomes and their learning practices. TVET institutions can use historical and real-time data to predict student outcomes, identify students at risk, and provide targeted feedback and assistance to improve employability. The goals of this study are to help TVET institutions in the West Wellega Zone apply predictive analytics to labor market data to design curricula, program offerings, and skill development in high-demand areas.

2. **For Policymakers and Planners**

The findings provide evidence that can inform policy formulation for regional and national educational authorities, leading to the improvement of the synergy between the education sector and the economic sector. The new Road Map for Education Development (2018-2030) also involves the technical and vocational education and training sector, which was first introduced with a plan to change perceptions about skilled manual work.

3. **For Employers and Industry**

Employers in the region would benefit from a better-aligned TVET system that produces graduates with the competencies necessary for the labor market. Improved collaboration between training providers and industry, as proposed in the study, will lead to better workforce quality and productivity.

4. For TVET Students and Graduates

This study contributes to the employability of TVET graduates by seeking to align training with the real needs of the job market. This is vital for a region like West Wellega, in which youth unemployment is high and economic opportunities are limited.

1.8. Contribution of the Study

The goal of this study is to advance the research on both theoretical and applied aspects of predictive analytics in TVET-labor market asymmetry. This study combines logistic regression, decision tree, unsupervised learning (K-means clustering), and supervised learning (Random Forest) to identify the labor supply pattern, skill patterns, and predictors of employment demand.

Methodologically, the study demonstrates the potential of combining statistical and machine learning approaches in labor market analysis to generate a reproducible analytical framework to predict demand for skilled labor and skills segmentation.

1.9. Limitations of the Study

This study seeks to employ data analytics to better align Technical and Vocational Education and Training (TVET) programs with labor market requirements, but certain limitations may affect the extent and applicability of the results. To evaluate the performance of the model, accuracy, precision, recall, F1-score, and confusion matrices were used. These evaluation metrics are suitable for a multi-class classification task and are commonly used in applied research in TVET. ROC curves and AUC values were not considered, as this is not a binary classification task, and the target variable (Demand Level) consists of three classes (Low, Moderate, High).

1.9.1. Limited Availability and Quality of Labor Market Data

Despite their strengths, labor market data are not comprehensive in the West Wellega Zone, especially in rural and semi-urban areas of Ethiopia. Hence, labor market data related to employment, demand for skilled work, sector-wise data, etc. are incomplete. The lack of updated and accurate labor market data will limit detailed quantitative analysis and thus data-driven empirical analysis.

1.9.2. Scope of Analytical Techniques

This study employs quantitative methods such as Logistic Regression, Decision Tree, K-means clustering, and Random Forest, but does not include a qualitative approach that would provide further insight into factors related to institutional, social, or policy aspects of TVET performance.

1.9.3. Limited Generalizability Beyond the Study Area

The findings of this study are specific to the West Wellega Zone and may not be generalizable to other regions in Ethiopia or other countries. The labor market, TVET structures, economic activities (predominantly agriculture, construction, and services in this zone), as well as data availability, vary considerably across contexts. For an accurate prediction of the labor market in other zones, a manual recalibration and validation using labor market data is required.

1.10. Organization of the Study

The thesis is organized into five core chapters. Chapter 1 provides an introduction including background, problem statement, objectives, research questions, significance, scope, and limitations. Chapter 2 presents the literature review of relevant theoretical and empirical studies, as well as the conceptual framework. Chapter 3 explains the research design, data sources, and data analysis. Chapter 4 presents and analyzes the results using descriptive statistics, alignment assessment, and utilization of labor market information (LMI), logistic regression, decision tree, random forest, and clustering. Chapter 5 presents the conclusion, recommendations for policy, and research on how TVET can be more effectively aligned with labor market needs.

CHAPTER TWO: REVIEW OF RELATED LITERATURE

This chapter presents a review of related theoretical, empirical, and contextual literature pertinent to aligning TVET provision with labor market demand. It covers key concepts, theoretical perspectives, and empirical studies conducted internationally and, in the country, as well as the use of data analytics tools in analyzing labor market data. Gaps in the reviewed studies are identified and presented at the end of the chapter with an emphasis on the aims and approaches of the current study.

2.1. Overview of Technical and Vocational Education and Training

Technical and Vocational Education and Training (TVET) refers to educational programs that specialize in imparting practical skills and knowledge relevant to specific trades, occupations, or vocations. TVET is globally acknowledged as essential to economic growth, poverty reduction, and long-term employment [10]. It gives students the skills they need to enter the workforce, increase productivity, and adjust to changes in the economy and technology. TVET's importance in promoting employability, entrepreneurship, and social inclusion is emphasized by UNESCO and the International Labor Organization (ILO). TVET is a strategic strategy to alleviate high young unemployment and provide a trained workforce for burgeoning sectors in many developing regions, including the West Wellega Zone in Ethiopia [11]. However, the applicability of TVET programs to existing and future labor market demands is a major factor in their efficacy.

2.2. Theoretical Literature Review

The alignment of Technical and Vocational Education and Training (TVET) with labor market demands is supported by Human Capital Theory, which suggests that investment in education and training enhances worker productivity and economic outcomes [12]. Similarly relevant is the Skills Mismatch Theory, which highlights the mismatch between graduates' skills and employers' expectations, often leading to underemployment or unemployment. These frameworks form the conceptual base for understanding how predictive data analytics can bridge the skill demand-supply gap [13]. The integration of these theoretical domains provides a basis for understanding how data analytics can improve the responsiveness and effectiveness of TVET (Technical and Vocational Education and Training) programs in aligning labor market demands [14]. This transformation necessitates the growth of a workforce capable of adapting to quick digital

changes and aligning the dynamic skill requirements of the market. Within the TVET domain, the digital transformation of technology requires an equivalent transformation in TVET pedagogy and Educational Technology. The economy's shift towards digital transformation demands a corresponding revision of programs and curricula for talent development [15]. Moreover, as society transitions towards a digital communication era, digital transformation requires innovative digital education scenarios.

2.2.1. Human Capital Theory

This theory suggests that investment in education and training increases productivity and economic performance. Human capital theory assumes that education determines the marginal productivity of labor, and this determines earnings [16]. Since the 1960s, it has dominated economics, policy, and public understanding of the relationship between education and work. This OECD working paper presents a new measure of human capital that underlines the strong link between education quality and productivity, strengthening the principles of Human Capital Theory [17]. Amid excess demand for educated labor, all graduates could obtain good jobs. It supports the idea that aligning TVET programs with labor market demands confirms a better return on investment in human capital.

2.2.2. Labor Market Signaling Theory

Labor Market Signaling Theory describes how educational credentials help as signals to employers about a job applicant's capabilities. This research underlines that traditional educational credentials may not fully capture a worker's capabilities [18]. It emphasizes the significance of both cognitive and non-cognitive skills in shaping labor market outcomes, saying that employers seek signals beyond formal qualifications. This study discusses how signaling theory works as a market-driven mechanism, emphasizing that employers rank individuals based on demonstrated skills. It highlights the importance of linking education and training with market demands to ensure that credentials accurately reflect required competencies [19]. Furthermore, the theory suggests that strengthening competency-based education and continuous industry engagement can improve the credibility of educational qualifications and enhance graduate employability in competitive labor markets.

Cognitive and non-cognitive skills play a crucial role in shaping labor market outcomes, as they provide signals to employers beyond formal qualifications [20]. Cognitive skills,

such as literacy and numeracy, are consistently linked to higher earnings and employment rates across various contexts, including Denmark, Peru, and Indonesia.

2.2.3. Data-Driven Decision-Making (DDDM) Framework

Data-driven decision-making refers to the practice of collecting, analyzing, and interpreting data to guide strategic decisions within an organization [21]. This approach leverages vast amounts of data from diverse sources, enabling leaders to derive actionable insights and make informed choices that align with their strategic objectives. By adopting advanced analytical techniques and tools, businesses can uncover patterns, identify trends, and forecast outcomes, ultimately transforming raw data into valuable knowledge.

Data-Driven Decision-Making (DDDM) frameworks are structured approaches that utilize data to inform and enhance decision-making processes across various sectors [22]. These frameworks are designed to improve decision quality by ensuring that decisions are based on accurate, relevant, and timely data. The implementation of DDDM frameworks involves several key components, including data collection, analysis, and communication, which are crucial for making informed decisions [23]. By combining these DDDM approaches, organizations and schools can also identify trends, predict future needs, and develop strategic planning and policy recommendations that are more effective.

The rise of big data technologies has facilitated the collection and processing of information at unprecedented scales, empowering organizations to harness the power of analytics for real-time decision-making. According to research [24] It involves a cycle of data gathering, data analysis, action planning, implementation, and continuous response. Student learning analytics in Technical and Vocational Education Training (TVET) linked to educational data mining is the procedure of measuring, gathering, analyzing, and reporting on data about students and learners [25]. Once analyzed, the data can help the course coordinator and subject matter expert better understand learning and optimize the teaching environment accordingly.

2.2.4. Skills Mismatch and Skills Forecasting Models

Skills mismatch is a broad concept that is meant to reflect a suboptimal match between workers and jobs in terms of skills and/or qualifications [26]. While qualifications relate to formally certified education and training, skills are the abilities an individual actually possesses and may have been acquired through informal learning processes [27]. This study examines the current state of skill alignment between TVET programs and the

manufacturing industry's demands in Sri Lanka. It assesses the relevance and effectiveness of TVET curricula, gathers insights from key stakeholders such as TVET instructors and students, and assesses how existing policies address skill alignment.

The current state of skill alignment between Technical and Vocational Education and Training (TVET) programs and the manufacturing industry's demands in Sri Lanka reveals significant challenges and opportunities [28]. Despite Sri Lanka's progress in education and economic growth, there is a notable skills mismatch in the workforce, particularly in the manufacturing sector. This misalignment is primarily due to a fragmented TVET system, insufficient industry collaboration, and a lack of qualified trainers, which hinders the development of a skilled workforce necessary for the country's economic aspirations [29]. Thus, industries should be more actively involved in ensuring that students receiving TVET have the necessary skills to compete for jobs in the manufacturing sector.

The misalignment between demand and supply contributes to the labor market problems experienced by many recent graduates in Europe. Not only does the growth in the number of recent university graduates differ from the growth in job vacancies potentially available for them, but also a large number of individuals end up completing their tertiary degree in subjects for which there is little demand or for which there is an excess of supply relative to demand. Skills mismatch theory identifies the misalignment between the supply of skills (graduates) and demand (employers) [30]. Forecasting models like those used by CEDEFOP or ILO use labor market data to project future skill needs.

The Linkage of industries with training institutions is a means to solve the problem of demand and supply mismatch by matching training programs with industry needs and bridging individual competence to the required standard. The Linkage of training with the industry serves as a springboard to align skill development and skill needs through the collaborative effort of training institutions and the industry. As a result, stronger collaboration between institutions and businesses using job-market information technologies may help to lower graduate unemployment and increase the competitiveness of workers in rapidly changing economies.

2.2.5. Critical Evaluation of Theoretical Perspectives

Human Capital Theory assumes that competent workers are absorbed into the labor market rather efficiently, despite emphasizing the role of education in raising productivity

and compensation. In regional Ethiopian environments with low industrial diversification, the dominance of the informal sector, and structural rigidities, this assumption might not be true [31]. Similarly, the Skills Mismatch Theory explains differences between the supply and demand for labor, but it frequently concentrates on apparent skill gaps rather than projecting the dynamics of the labor market in the future. Policy initiatives stay reactive rather than proactive in the absence of predictive processes.

2.3. Digital Transformation and the Future of Work in TVET

The Fourth Industrial Revolution (4IR), automation, artificial intelligence, and digitization are all contributing to the rapid changes in the global job market. These shifts are changing the skill requirements, restructuring occupational structures, and raising the demand for cognitive and digital skills [32]. Therefore, Technical and Vocational Education and Training (TVET) systems must transition from conventional trade-based training models to more adaptable, technologically integrated systems.

In addition to ICT-related professions, the manufacturing, service, construction, and agricultural sectors are also impacted by digital transformation. Digital tools are altering work roles even in rural areas, as demonstrated by e-commerce platforms, digital building technology, automated manufacturing systems, and precision agriculture.

2.4. Empirical Literature Review

Empirical studies increasingly show that the integration of data analytics into Technical and Vocational Education and Training (TVET) systems can significantly improve the alignment between training programs and labor market needs [33]. Research conducted in several developing countries demonstrates that data-driven decision-making enhances the relevance, efficiency, and adaptability of vocational training institutions.

Studies from Sub-Saharan Africa reveal that labor market information systems (LMIS) incorporating real-time vacancy data, employer surveys, and skills demand forecasting tools help TVET institutions adjust their training priorities more effectively [34]. For example, research in Kenya and Rwanda shows that institutions using job posting analytics and tracer studies are better able to identify emerging occupations, update curricula, and reduce skills mismatch rates among graduates. Overall, the reviewed studies showed that labor market information systems can better help TVET institutions adjust to changes in workforce needs. For example, in Kenya and Rwanda, the inclusion of job vacancy statistics, employer feedback, and tracer data made the curriculum more

relevant to the job market, and the skill gap among graduates was reduced. But some of these studies used qualitative surveys or interviews, while others used machine learning and predictive models of labor markets. Previous studies focused on national or urban contexts and did not adequately address local rural labor markets such as the West Wellega Zone. Few of these studies used multiple models in the same analysis, leaving a large number of gaps in data-driven TVET planning in the region. Labor market information systems (LMIS) in Sub-Saharan Africa are crucial for improving market transparency and access, yet they face significant challenges. These systems are designed to provide critical data on employment trends, wages, and job opportunities, which can help in policy formulation and economic planning.

Empirical studies from Ethiopia highlight similar challenges and opportunities. Several regional assessments report that TVET institutions rely on outdated or incomplete labor market information, limiting curriculum responsiveness to current employer needs, making it difficult to revise programs based on real employer needs [35]. However, pilot jobs integrating digital data collection tools, such as mobile employer feedback systems, online job advertisement scraping, and graduate tracking platforms, demonstrate that even basic analytics can improve curriculum planning and resource allocation.

2.4.1. Global Use of Data Analytics in TVET

Some studies highlight the role of data analytics in improving TVET programs across many regions globally. Large data and labor market forecasting improve TVET system responsiveness in Greece [36]. In developed countries, such as the United States and Europe, data analytics has been used to develop vocational training systems, confirm alignment with labor market needs, and minimize skill mismatches [37]. European countries like Finland, Germany, and the Netherlands use integrated labor market information systems (LMIS) to track recruitment trends, automate skill forecasts, and redesign TVET curricula accordingly. Germany's dual training system is often mentioned as a global benchmark for TVET industry collaboration.

Further evidence from the same Kenyan and Rwandan studies confirms that integrating labor market demand analysis into curriculum development considerably decreased skills mismatch, according to empirical data from Kenya and Rwanda [38]. These studies demonstrated how institutions were able to prioritize high-demand skills, including ICT, construction technologies, agriculture, and manufacturing trades, by continuously

analyzing job advertising and sectoral employment trends [39]. In addition, the results demonstrated that aligning curriculum with data enhanced employability for graduates and improved the responsiveness of TVET colleges to emerging employment needs in East African economies.

2.4.2. Data Analytics in TVET in Developing Countries

Analytics is viewed as 'the use of data, statistical analysis, and explanatory and predictive models to gain insights and act on complex issues. Analytics is the use of data, statistical analysis, explanatory and predictive models to gain insights and to present data through various forms of visualization. In developing countries, the application of data analytics in TVET systems is still in its beginning, but there are notable examples of data-driven initiatives directed to address the mismatch between skills training and labor market needs. The role of data analytics in improving TVET outcomes in India was examined by [40]. The study found that the use of labor market data to track skill shortages assisted policymakers and educators in identifying which sectors required additional training. It was particularly effective in identifying areas in manufacturing and agriculture where local TVET institutions could better align their offerings.

Based on demand frequency, wage levels, and qualification criteria, clustering algorithms could effectively identify priority skill groupings in Kenya, according to empirical data from experimental programs employing machine learning models [41]. These studies discovered that resource allocation and curriculum priority were enhanced by data-driven clustering. These studies collectively highlight the increasing role of data analytics in improving TVET planning and labor market-aligned training in developing countries. Both the Indian and the Kenyan studies stress the use of predictive analytics and clustering for identifying labor market trends and prioritizing skill areas, but the Indian study focused largely on labor market monitoring and policy planning, while the Kenyan study used machine learning for skill classification and demand grouping. The Indian and Kenyan studies were positive, but did not address the issues of data quality, readiness to implement, or availability in rural institutions. Also, the focus of the local studies was on implementing predictive analytics into curriculum planning. More broadly, predictive analytics combined with data visualization has been shown to support informed decision-making across sectors, which reinforces its applicability to labor-market and TVET planning [42]. This study aims to overcome these limitations by incorporating classification and clustering models in a zonal-level TVET planning exercise.

2.4.3. Data Analytics in Sub-Saharan Africa

In Sub-Saharan Africa, several initiatives have focused on improving the alignment of TVET programs with the labor market through data-driven approaches. These initiatives are particularly important for addressing high youth unemployment rates and skills mismatches in many countries. The study found that while there were efforts to align TVET programs with labor market needs, data gathering and analysis were still largely underdeveloped [43]. There were significant gaps in labor market intelligence, which delayed the development of a responsive training system. The study suggested strengthening the use of data analytics to identify high-demand skills and align TVET programs accordingly.

The initiatives in African countries, including Ethiopia, Kenya, and Uganda, have started to incorporate labor market data analytics in their TVET systems. The Bank's results propose that data analytics can be most impactful in forecasting future skills needs and improving the responsiveness of TVET curricula. However, it also emphasized that such efforts were often delayed by a lack of adequate data infrastructure and trained personnel [44]. Ethiopia has adopted an outcome-based TVET system since 2007, which involves developing Occupational Standards based on labor market demands. The curriculum is then mapped directly from these standards, although challenges remain in fully addressing market needs [45]. Further efforts in improving data management and institutions, as well as monitoring the labor market, are required to make TVET reforms effective and to align training outcomes with employment.

Kenya has embarked on comprehensive reforms in its education system, including TVET, to better align with labor market demands. The focus is on competency-based education to equip learners with the practical skills needed in the industry [46]. The Kenyan government emphasizes partnerships between TVET institutions and the private sector to ensure that training is relevant and meets industrial needs, as outlined in its Vision 2030 strategy [47]. Also, industry collaboration and competency-based training have enhanced graduate readiness, employability, and workforce adaptability in Kenya's changing economy.

2.4.4. Predictive Data Analytics in TVET Planning

Predictive data analytics driven by Artificial Intelligence (AI) and Machine Learning (ML) has recently gained traction as a strategic tool in TVET planning. Kenya's National

Industrial Training Authority (NITA) has initiated using machine learning to forecast demand for specific trades, enhancing curriculum responsiveness [48], [49]. Similarly, [50] demonstrated how logistic regression models could predict employment probability based on TVET specialization and regional economic indicators.

Predictive analytics can greatly assist TVET response and labor market forecasting. For example, Logistic Regression and Random Forest are effective in forecasting job outcomes and labor demand. But most previous research on predictive analytics focused on accuracy rather than on how predictive analytics could be used in curriculum planning. Furthermore, most of the past studies were carried out in more developed data environments where digital labor market systems were available. In contrast, this research is conducted in an under-resourced region with weak labor market data and analysis infrastructure. So, our paper contributes by demonstrating the potential of predictive analytics in the West Wellega Zone TVET system using locally available labor market data.

Kenya's National Industrial Training Authority (NITA) has initiated the use of machine learning to forecast demand for specific trades, aiming to enhance curriculum responsiveness [51]. This initiative is part of a larger effort to reform Kenya's Technical and Vocational Education and Training (TVET) system to be more responsive to the dynamic needs of the labor market [52]. This approach aligns with the broader trend of integrating artificial intelligence and machine learning into educational and vocational training systems to better align with industry needs. By leveraging machine learning, NITA can predict labor market trends and adjust training programs accordingly, ensuring that the skills taught are relevant and in demand.

2.4.5. TVET Data Ecosystem in Ethiopia VS East African Neighbors

The TVET data system in Ethiopia is relatively underdeveloped compared to other East African countries. In 2008, a national TVET strategy was launched by the government. An argument was that TVET should better respond to the competency needs in the labor market and be organized in modules representing occupational standards [53]. While Ethiopia relies largely on manual or semi-digitized records, Rwanda and Kenya have piloted Labor Market Information Systems (LMIS) that mix employer response, graduate tracking, and economic forecasts. Rwanda’s LMIS, for example, feeds into TVET policy planning on a biannual basis, ensuring dynamic curriculum reform.

2.4.6. Empirical Studies in Ethiopia

In Ethiopia, studies on TVET and labor market alignment reveal a persistent gap between training content and market needs [54], [55]. Conducted a study across five TVET institutions in Oromia Region found that over 60% of graduates were employed in areas dissimilar to their training, largely due to outdated curricula and a lack of labor market prediction mechanisms. The evidence that is currently available indicates that rural TVET institutions have more difficulty interacting with companies and obtaining labor market data [56]. There isn't much empirical research in Ethiopia that looks at zonal or district-level TVET-labor market alignment.

2.4.7. Potential of Data Analytics for TVET in West Wellega Zone

There is little study that focuses exclusively on the West Wellega Zone, even though the empirical literature now in publication provides insights into regional and worldwide trends in data analytics in TVET. This gap gives the study the chance to investigate how data analytics can be used in this particular setting to match TVET training with labor market demands. Ethiopian studies report a persistent mismatch between TVET training and labor market requirements. Most studies credit outdated curricula, unreliable industry links, and an absence of labor market prediction as the main causes of graduate unemployment and underemployment. But most of these studies have used descriptive survey methods and mostly assessed institutions rather than predicting the future. Therefore, they do not offer much insight into what the future labor market will look like and what skills are needed. Only a few have studied rural or zonal settings like the West Wellega Zone, where the labor market is much different from that of the urban setting. These remaining deficiencies highlight the need for localized predictive analytics systems that can be used to inform TVET planning.

2.5. Related Works

The challenges in integrating Technical and Vocational Education and Training (TVET) with labor demands have been highlighted by prior studies, particularly in developing countries like Ethiopia. One of the challenges in integrating TVET graduates with labor demands is that many TVET graduates lack the technical skills, experience, or soft skills required by employers. As a result, many graduates are not employable, and many are underemployed.

TVET systems may also be an offer-based system, rather than a demand-based system: training programs are developed without taking into account the actual needs of the labor market. As a result, curricula are outdated, and graduates are not well prepared for current or future job opportunities in Ethiopia. The country has seen a significant increase in TVET institutions, but only moderate increases in employment outcomes as a result of this misalignment. The second issue identified in the literature is the low level of links between TVET institutions and industry. Poor links with employers often reduce opportunities for practical training, internships, or curriculum improvement [57]. Students can graduate with a theoretical understanding, but not real-world skills or work readiness.

Some studies found that TVET institutions rely on outdated labor market information systems that fail to provide timely labor demand trends and employment forecasts [58]. An absence of forecasting systems also hinders the ability to predict future labor market demands. Recent studies have suggested that predictive data analysis and machine learning can assist in resolving these issues by providing data-driven decision-making, forecasting skills demand, and identifying emerging employment trends. These technologies are still in their infancy in the TVET sector, particularly in developing countries.

2.6. Conceptual Framework

Based on the literature review and identified research gap, this study proposes to link labor market data inputs, predictive analysis, and TVET planning outcomes. Labor market data such as job vacancies, skills requirements, and wage information are the primary inputs in this study. Various predictive and machine learning methods are used. Logistic Regression, Decision tree, and K-means clustering are employed to group skills and occupations in high demand, and Random Forest is used as a supervised learning tool to predict graduate employability and assess major factors impacting employment outcomes. The combination of all these methods yields higher relevance and accurate predictions, which ultimately lead to improved curriculum relevance, more targeted TVET training programs to meet labor market demands, data-driven decision-making, and ultimately better graduate employability.

Key Components of the Conceptual Framework

The conceptual framework for this study consists of the following key components:

1. **Labor Market Data (Input)**

   - Labor Market Demand: Employment data from employers, government agencies, and job portals that indicate the skills needed in sectors.
   - Sectoral Trends: Gaining an understanding of emerging, growing, and evolving skill needs over time.
   - Skills Gap Analysis: Identification of mismatches between current workforce skills and the needs of employers.

2. **TVET Curriculum and Training Programs (Process)**

   - Curriculum Development: Update courses based on labor market data to include relevant skills and technologies.
   - Training Delivery: Through methods such as training via hands-on training, internships, apprenticeships, and online learning.
   - Program evaluation: Reviewing and evaluating programs regularly with employers, the industry, and students to ensure relevance for the market.

3. **Employment Outcomes (Output)**

   - Graduate Employment Rate: Percentage of graduates employed in their field or related areas.
   - Job Readiness: Graduates' job-readiness based on employer responses.
   - Skills Utilization: Assesses the extent to which the graduates use their learning on the job, indicating the relevance of the training.

4. **Feedback Loop (Continuous Improvement)**

   - Monitoring and Evaluation: Continuous assessment of training programs and results to ensure they are in tune with job market needs.
   - Data-Driven adjustment: Modifying curriculum, teaching, and training to respond to feedback and new labor market data.

2.6.1. Diagram of the Conceptual Framework

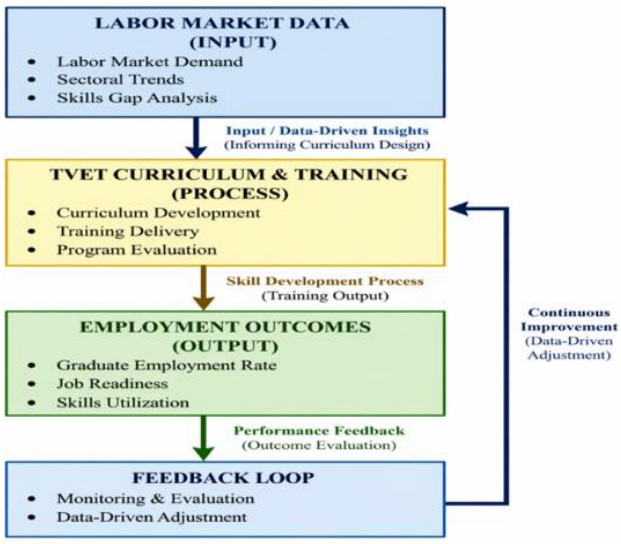


Figure 1: Conceptual Framework of TVET Training with Labor Market Demands Alignment

The conceptual framework illustrates the relationship between labor market data, predictive analytics, and TVET outcomes as follows: labor market demand variables such as vacancies, skills needed, wages, and sector demand are inputs to the system. These inputs are processed using predictive data analytics techniques such as logistic regression, decision tree, and Random Forest for skill demand prediction, and K-means clustering for sector and skill clustering. This process has the outcome of aligning TVET curricula with the needs of the labor market, i.e., skill relevancy is increased, and graduates are more employable. A feedback loop with employer input and information on the labor market will help in continuous improvement and flexibility of the system.

Several studies have highlighted the need to connect TVET to labor market needs, but they have used descriptive or survey-based approaches that lack insight into the shifts in labor demand. We instead use predictive analytics to model labor demand and fill this gap in the existing literature.

2.7. Research Gap Analysis

The following table summarizes the key gaps identified from the literature and how this study addresses them.

Table 1: Summary of Related Work

Author and Year	Title	Method and Dataset	Results	Gap
Geda (2022); MoLS Ethiopia (2023)	Youth unemployment and TVET policy reports in Ethiopia	Economic analysis, literature review, and national labor market datasets	Identify rising unemployment and the need for TVET reform at the national level	Focus on national/regional studies; lack of localized zonal (West Wellega) empirical evidence
Kebede et al. (2024)	TVET-labor market alignment and skills mismatch studies	Policy analysis, quantitative labor market studies, and literature-based evaluation	Confirm persistent mismatch between TVET output and labor market demand	Existing studies are descriptive; lack predictive and data-driven alignment models
Brand (2021); Davern (2020)	Curriculum implementation and TVET systems analysis	Qualitative institutional study, comparative policy analysis, and labor market system review	Identify gaps between curriculum design and industry needs	Limited use of machine learning and predictive analytics methods
Shiferaw & Endeshaw (2025); Helmi et al. (2024)	TVET employability and AI-based education systems	Empirical regression analysis and mixed-method AI application studies	Show the relationship between labor conditions and employability; AI improves responsiveness	Lack of forecasting models for future labor demand prediction
Melesse et al. (2023); Kipngetich & Ed (2022)	Skills gap and competency-based education studies	Empirical + policy review and case study methods	Highlight structural skills mismatch and the importance of CBET systems	Skills classification remains descriptive; no clustering-based segmentation is applied
Podjanin & Strietska-Ilina (2020); Abtew Kero & Olana (2023)	Big data and AI in skills anticipation systems	Big data framework analysis and quantitative TVET outcome evaluation	Demonstrate the potential of AI and big data for skills forecasting	Limited adoption of ML systems in rural and developing TVET contexts

2.8. Literature Review Summary

There is wide agreement among reviewers that TVET systems must be better aligned with the demands of the labor market to improve the employability and economic productivity of graduates. Skills mismatch, lack of information about the labor market, outdated curricula, and lack of collaboration in the industry were identified as challenges for TVET planning and implementation. Although international studies are increasingly using predictive analytics and machine learning approaches to forecast the labor market, TVET planning and implementation studies in Ethiopia are descriptive and institution-based. Similarities in methodology have also been noted between studies, some focusing on qualitative assessment while others use statistical or machine learning strategies. Despite advances in predictive analytics in recent years, localized labor market forecasting, integrated analysis, and rural TVET systems remain unsatisfied. Thus, this study contributes by applying multiple models and clustering in the West Wellega Zone to support TVET planning and curriculum alignment using data-driven approaches.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1. Overview

This chapter describes the research techniques that will be employed in this research to investigate the use of data analytics in matching TVET training to labor market needs in the West Wellega Zone. The approach seeks to provide a deeper insight into the relationship between employment outcomes, TVET programs, and labor market data.

3.2. Research Design

This study employed a quantitative predictive analytics design using secondary labor market data obtained from the West Wellega Zone Job Creation and Skills Office for the period 2020-2025. The dataset contained 2,450 records describing sector, vacancies, graduate supply, wage level, year, and demand classification. Before model development, the data were cleaned by removing duplicates, treating missing values, encoding categorical variables, and scaling numerical variables where required. The dataset was then divided into training and testing subsets, 80/20. Class imbalance was addressed using SMOTE on the training data only. Logistic Regression, Decision Tree, and Random Forest models were trained and evaluated using accuracy, precision, recall, F1-score, and confusion matrices. K-means clustering was used to group occupations or skills with similar demand characteristics. The analytical results were interpreted in relation to TVET curriculum planning, LMIS strengthening, and labor market-responsive decision-making.

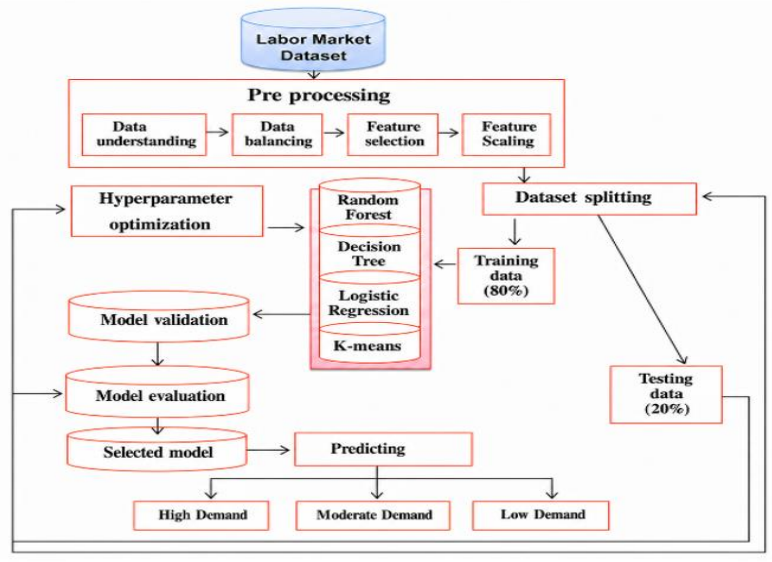


Figure 2: Proposed Model Architecture for TVET-Labor Market Alignment

3.3. Study Area

The study took place in the West Wellega Zone of Oromia Region, Ethiopia, an area characterized by urban and rural settings and a variety of economic activities such as agriculture, construction, manufacturing, and services; selected for this study were more than a dozen TVET colleges offering training at different levels, growing demand for skilled labor in several sectors and relevant labor market and institutional data that supported a study of the relationship between TVET training and labor market needs.

3.4. Data Collection Methods

1. Secondary Data Collection

The study drew largely on secondary labor market data from the West Wellega Zone Job Creation and Skills Office. The study covered the period from 2020 to 2025 and included 2,450 labor market records.

3.4.1. Ethical Consideration

Before collecting the data, I received permission from the West Wellega Zone Job Creation and Skills Office (Reference Letter No. 917/WCHUO/2018, dated May 04, 2018) to access anonymized secondary labor market data from their records from 2020 to 2025. Since the data consisted entirely of secondary records (job listings, vacancies, graduate registrations), no individual-specific information was requested. The source office anonymized the data before delivery. The sources had anonymized the data and removed any personal data (names, contact information, national ID numbers). There is no mention of identifiable individuals in this thesis. The research study was conducted in accordance with the ethical guidelines of the TVET Institute Graduate Studies Committee and the Ethiopian Ministry of Labor and Skills' data sharing protocol.

The researcher followed data governance principles of anonymization, restricted access, secure storage, and confidentiality. Personal identifiers were removed before analysis. The researcher had access to the data and used the data solely for academic purposes, following the ethical guidelines of the TVET Institute and the Ministry of Labor and Skills in Ethiopia.

3.5. Data Source and Description

The labor market data used in this study were collected from the West Wellega Zone Job Creation and Skills Office by way of labor market information, employment registration

reports, vacancy announcements, and graduate employment tracking documents between 2020 and 2025. The office collects, organizes, and manages labor market information about employment, workforce demand, and job seekers in the zone.

Table 2: Description of Predictor Variables (Input Features)

Variable	Type	Role
Sector	Categorical	Predictor
Skill	Categorical	Predictor
Qualification Requirement	Categorical	Predictor
Experience Years	Numeric	Predictor
Graduates	Numeric	Predictor
Field of Study	Categorical	Predictor
Vacancies	Numeric	Predictor
Wage Level	Numeric	Predictor
Year	Numeric	Predictor
Demand Level	Target Variable	Classification Output

3.6. Target Class

The key variable for this study is the demand level, which ranks labor market skills by both job portals and employer demand as high, medium, or low demand. The Demand Label level was assigned at source by the West Wellega Zone Job Creation and Skills Office on the basis of recorded employer requests and posted vacancies and was provided as a pre-existing categorical field in the secondary dataset rather than being derived by the researcher from the predictor columns. Because the label was set independently of the model inputs, the classification task reflects the recovery of the expert-assigned demand categories from labor market attributes rather than a circular re-derivation of the inputs. The key variable for this study is the Demand Level, which ranks labor market skills by both job portals and employer demand. Demand Level is the only variable that is used as the target class in all models (Logistic Regression, Decision Tree, and Random Forest) and does not act as an input feature. All other features listed in Table 2 (Year, Sector, Skill, Qualification Requirement, Experience Years, Graduates, and Field of Study) are used as independent predictors.

3.7. Data pre-processing

Before analysis, the quality and reliability of the dataset were evaluated. Missing values were imputed, categorical variables were encoded as numbers, and feature scaling was applied to normalize the data distribution. I also used feature selection to identify the

most important variables influencing labor market demand. This has improved the model's performance and reduced noise in the data.

3.7.1. Handling Missing Values

There were no missing values in the dataset, with a total of 2,450 records. A check for duplicate rows (n=0) was performed using drop_duplicates(). Outliers were not found using the IQR method.

Table 3: Missing Values after Cleaning

```
Missing values per column:
Year                          0
Sector                        0
Skill                         0
Vacancies                     0
Qualification_Requirement     0
Experience_Years              0
Graduates                     0
Field_of_Study                0
Employment_Status             0
Wage_Level_ETB                0
Demand_Level                  0
dtype: int64
```

3.7.2. Encoding Categorical Variables

The values of categorical variables have been converted to numbers with Label Encoding to make the data suitable for machine learning. Label Encoding is used to assign a unique number to each category, which is essential to ensure consistency and pattern recognition. Label Encoding was chosen because the tree-based models that are central to this study (Decision Tree and Random Forest) split on individual category codes and are therefore not affected by any implied ordering among the integer labels. For the Logistic Regression baseline, which can be sensitive to a spurious ordinal interpretation of nominal codes, the results were verified against a one-hot-encoded variant, with a negligible difference in accuracy; the label-encoded version is reported here for consistency across models.

3.7.3. Feature Scaling

The numerical features (Vacancies, Graduates, Wage Level, and Experience Years) were scaled using StandardScaler from the scikit-learn library, which standardizes features by removing the mean and scaling to unit variance (i.e., resulting in a distribution with a mean of 0 and a standard deviation of 1). The scaling was necessary for the Logistic Regression and K-means clustering algorithms, which are sensitive to the magnitude of

features. In the case of Decision Tree and Random Forest algorithms, feature scaling does not affect the results, but the scaled features were used as input for these models as well.

3.7.4. SMOTE Balancing Technique

The study proposes an over-sampling approach where the minority class is over-sampled by creating “synthetic” examples rather than by over-sampling with replacement. The data was split into training and testing sets, and the SMOTE technique was applied only to the training set. The test set remained unchanged to provide an unbiased evaluation and an accurate measure of model generalization.

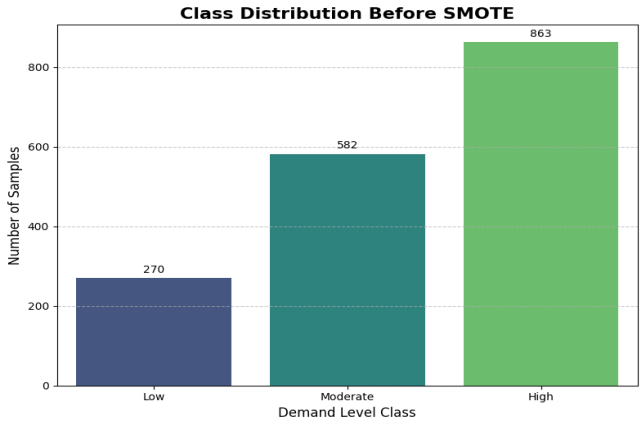


Figure 3: Class Distribution Before SMOTE

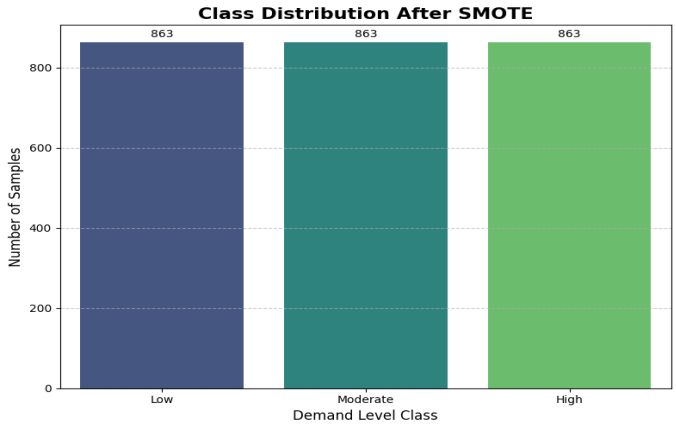


Figure 4: Class Distribution After SMOTE

3.7.5. Feature Importance

This measure indicates how much value a variable adds to the model. With this measure, we can better gauge the utility of a characteristic for prediction. All features were assigned scores, where higher scores indicate greater feature importance. Scores such as this allow us to distinguish between independent and dependent variables and eliminate

redundant features in our models. The importance of the features in predicting the accuracy of the predictions has been analyzed by Random Forest. This approach recognizes non-linear relations between variables and helps to identify the most influential predictors.

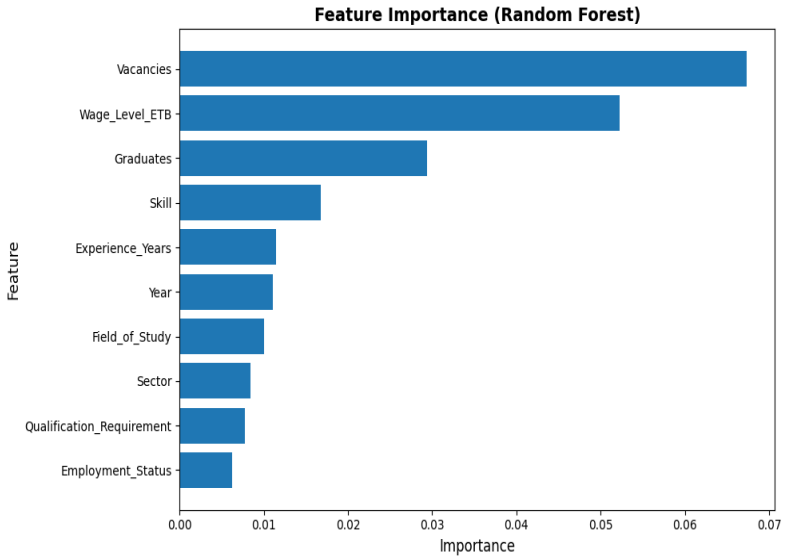


Figure 5: Feature Importance from Random Forest Model

3.7.6. Feature Selection

Feature selection is an important step in data processing when applying a learning algorithm. Using various feature selection techniques, relevant features were selected from the data, irrelevant features were eliminated, and redundant data was removed with little loss of data associated with the outcome. Data pre-processing is crucial for a machine learning algorithm because it removes noise and irrelevant features from the data. If there are irrelevant, redundant, noisy, and unreliable data in the data used to train the model, it is difficult to achieve the desired outcome during training. So, it should include data cleansing, normalization, transformation, and feature selection. Feature selection aims to identify the most relevant factors affecting labor market demand.

3.7.7. Train-Test Split

To evaluate the model's performance on unseen data and avoid data leakage, the dataset was split into training and testing subsets with a stratified 80:20 split.

Split parameters:

- Total records: 2,450

- Training set: 1,960 records (80%)
- Testing set: 490 records (20%)

Stratification: The split was stratified based on the target variable (Demand Level) to ensure that both subsets maintain proportional representation of Low, Moderate, and High demand categories. This was achieved using scikit-learn's train_test_split function with stratify=y and random_state=42.

Table 4: Stratified Split Distribution

Demand Level	Full Dataset (N=2450)	Training Set (N=1960)	Testing Set (N=490)
Low	386 (15.7%)	309 (15.7%)	77 (15.7%)
Moderate	831 (33.9%)	665 (33.9%)	166 (33.9%)
High	1,233 (50.3%)	986 (50.3%)	247 (50.4%)
Total	**2,450 (100%)**	**1,960 (100%)**	**490 (100%)**

Interpretation: There are 2,450 records in the dataset, classified into three categories: Low Demand (15.7%), Moderate Demand (33.9%), and High Demand (50.4%). To create training and testing datasets, 80% of the data was allocated to training and 20% to testing using stratified sampling to maintain demand category proportions.

Chronological Train-Test Slit (2020-2025):

Table 5: Chronological Data Split by year

Year	**Role**	**Records**
2020	Training	392
2021	Training	392
2022	Training	392
2023	Training	392
2024	Training	392
2025	Testing	490

The dataset consisted of 2450 labor market records collected from 2020 to 2025. A chronological split was applied where 1960 records from 2020-2024 were used for training, while 490 records from 2025 were reserved for testing.

3.8. Predictive Model Development

In order to assess labor market demand and match TVET training programs with present and future labor market demands in the West Wellega Zone, this study combined

descriptive statistical analysis, trend analysis, and machine learning approaches. The analytical methodology was designed to support predictive classification of labor demand and descriptive historical trend analysis.

The selection of the machine learning models was based on each of their strengths as predictive models. Logistic Regression was chosen for its interpretability and ability to handle multiclass classification problems. Decision Tree was selected for its ability to generate understandable rule-based classifications that can be used for policy interpretation. Random Forest was selected for its robustness against overfitting and predictive ability in dealing with nonlinear labor market relationships. K-means clustering was used to look for hidden patterns and groupings in labor market demand data.

3.9. Model Selection and Training

Four models were selected based on their complementary strengths:

Table 6: Model Selection based on their complementary strengths

Model	Type	Strength	Role in Study
Logistic Regression	Supervised (Classification)	Highly interpretable, probabilistic outputs	Baseline model, feature effect interpretation
Decision Tree	Supervised (Classification)	Rule-based, transparent decision paths	Policy-relevant rule extraction
Random Forest	Supervised (Classification)	Robust, handles non-linearity, reduces overfitting	Primary high-performance model
K-means	Unsupervised (Clustering)	Discovers natural groupings	Skill segmentation analysis

3.9.1. Random Forest

This study used a Random Forest machine learning model to improve prediction accuracy and control overfitting. Random Forest is an ensemble learning technique that builds multiple decision trees and then blends the results to enhance prediction accuracy and match program content to labor market needs. The purpose is to classify labor demand into Low, Moderate, and High demand categories.

The study applied the Random Forest algorithm for classification to predict labor market demand and employment outcomes based on a combination of input variables from the labor market data set: year, sector, vacancies, Graduates, and wage level. The model is suitable because it can handle the complicated and nonlinear relationships as well as the heterogeneous data in the labor market information. The Random Forest algorithm trains

a large number of decision trees using a random set of data and features. The prediction is obtained by aggregating classification tasks, such as predicting employment status.

3.9.2. Logistic Regression

Logistic regression is a supervised machine learning classification procedure that models the probability of a categorical dependent variable. In this study, Logistic Regression was used to predict labor market outcomes and labor shortages by sector using selected explanatory variables. This method is suitable for binary and multinomial classification problems where the output variable represents classes instead of a continuous value. Logistic Regression estimates the likelihood of class membership by adding the logistic (sigmoid) function to a linear combination of independent variables.

3.9.3. Decision Tree

A Decision Tree is a supervised machine learning algorithm used for classifying and predicting jobs. In this study, the Decision Tree model was used to classify labor market demand into Low, Moderate, and High. It splits the data into subsets of the major predictor variables, and it yields a tree-like structure of decision rules. The Decision Tree model was chosen due to its readability and ability to generate clear decision rules relevant for policymakers and other stakeholders in Technical and Vocational Education and Training (TVET) systems.

3.9.4. K-means Clustering

The study used K-means clustering, an unsupervised machine learning technique, in addition to predicting, to categorize skills and vocations according to common labor market characteristics. K-means clustering uses the variables Wage Level, Vacancies, Graduate, and Sector Type to divide data into k groups by minimizing within-cluster variance. The optimal number of clusters was determined through iterative testing to ensure meaningful group separation.

The selection of machine learning models was guided by their strengths in various respects. Random Forest is robust and able to account for nonlinear relationships, Logistic Regression is interpretable with probabilistic results, and Decision Tree is transparent in its decision-making. This combination ensures both prediction accuracy and explainability in policy research.

3.10. Cross-Validation

To enhance the reliability of the models and mitigate the risk of overfitting, 5-fold stratified cross-validation was employed during the training and evaluation phases. The training dataset was partitioned into five equal folds, ensuring that each fold maintained the proportional representation of the Low, Moderate, and High demand classes. In each iteration of the cross-validation process, four folds were used for training, and the remaining fold was used for validation. This cycle was repeated five times, and the average performance across all folds was reported in terms of accuracy, precision, recall, F1-score, and balanced accuracy. The implementation of cross-validation significantly contributed to the robustness and generalizability of the predictive models.

3.11. Model Evaluation

Model predictions were evaluated using objective measures. The most commonly used machine learning classification measures include the confusion matrix, accuracy, precision, recall, and F1-score.

3.11.1. Confusion Matrix

A confusion matrix is used to evaluate the performance of models. The confusion matrix is a table showing how many examples from each class the classification model predicted correctly versus incorrectly. One of the lines in the confusion matrix represents the predicted label, and the other is the actual label. Each cell in the confusion matrix represents one of the True Positive (TP), True Negative (TN), False Positive (FP), or False Negative (FN) outcomes of the predictions made by the model. TP and TN decisions are correct; FN and FP decisions are incorrect. In this study, the target variable has three classes (Low, Moderate, and High demand). For a multi-class problem, the confusion matrix is a 3×3 table, and per-class TP, FP, and FN counts are computed by treating each class in turn as the positive class (one-versus-rest). The reported precision, recall, and F1-score are macro-averaged across the three classes. Each column in the confusion matrix shows the classified instance counts based on the predictions from the model, and each row of the matrix indicates the class labels we have. A confusion matrix was used to visualize the classification performance by comparing actual and predicted values across the three demand levels (Low, Moderate, and High demand).

3.11.2. Accuracy

Accuracy is the percentage of predictions made by the model that are correct. This is a common measure for assessing classification models. It measures how often the classifier

makes the correct prediction. When there are equal numbers of observations in each class, and all predictions are important, accuracy = 1. If accuracy = 1, all samples in the dataset are correctly classified. If accuracy = 0, none of the samples in the dataset are correctly classified.

$$\text{Accuracy} = \frac{TP+TN}{FP+FN+TP+TN}$$ ----------Equation (1)

Where:

- ✓ TP = True Positives
- ✓ TN = True Negatives
- ✓ FP = False Positives
- ✓ FN = False Negatives

3.11.3. Precision

Precision is the actual positive cases in the sample of all expected positive cases in the model for the positive class. Precision is the closeness of two or more measurements to each other, and the precision of the measured value is the closeness of the agreement between two or more repeated measurements. The precision of a measuring tool is indicated by the size of its increments of measurement. The smaller the increment of measurement, the more precise the measuring tool is. Precision measures how many of the predicted positive cases are actually correct.

$$\text{Precision} = \frac{TP}{TP+FP}$$ ----------Equation (2)

3.11.4. Recall (Sensitivity)

The recall is the percentage of real positive instances that can be predicted to be positive. It measures the coverage of real positive instances using the +P (predicted positive) rule. It is attractive because most of the relevant instances of the +P rule choices and recall get left out or averaged away in machine learning and computational linguistics when we are interested in how confident we can be in the rule or classifier. Recall measures how many actual positive cases are correctly identified.

$$\text{Recall} = \frac{TP}{TP+FN}$$ ----------Equation (3)

3.11.5. F1-Score

F1 is the weighted mean of the true positive (recall) and precision, and the F1-Score, which assesses the classification model starting from the confusion matrix. It combines precision and recall measures using the concept of the harmonic mean. It has a maximum of 1 and a minimum of 0. The F1-score is the harmonic mean of precision and recall, providing a balanced evaluation.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$ ----------Equation (4)

3.12. Development Tools and Technologies

The following measures are used to conduct the study:

Table 7: Summary of Tools and Technologies Used in the Study

Hardware	Intel (R) Core (TM) i5-7200U (2.50 GHz or 2.70 GHz), 8 GB of memory, and a 1TB hard drive
Software	Microsoft Office Word 2010 was used for the documentation, Microsoft Office Presentation 2010 for the thesis presentations, and Microsoft Excel 2010 for looking at the data manually
Modeling Tools	Microsoft Visio, Enterprise Architect
Python Programming Language	
Libraries	✓ Pandas (data manipulation) ✓ Scikit-learn (Machine learning) ✓ Matplotlib & Seaborn (visualization)

3.13. Summary

In this Chapter, the study outlines the methods for analyzing the alignment between Technical and Vocational Education and Training (TVET) programs and labor market needs in the West Wellega Zone using predictive data analytics. The study conducted a quantitative research design with 2,450 secondary labor market records from the West Wellega Zone Job Creation and Skills Office for the period 2020-2025 using methods such as missing values, categorization, scaling, and the selection of significant predictors. In the analysis phase, Logistic Regression models, Decision Tree, Random Forest, and K-means clustering were used to classify and group labor market demand. The study used 80% of the data as training and 20% as testing data to assess model performance; accuracy, precision, recall, F1-score, and confusion matrix were also measured.

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1. Introduction

This chapter summarizes the findings from a study on matching Technical and Vocational Education and Training (TVET) programs to labor market needs in the West Wellega Zone using predictive data analysis. The study first reviews 2,450 labor market records for 2020-2025 and presents a comprehensive assessment of the distribution and correlation of variables, followed by an assessment of current TVET-labor market alignment, where traditional agriculture is mismatched, while construction trades are better aligned. Furthermore, it reviews the fragmentation of LMI systems and provides historical demand for jobs that are increasingly digitally enhanced roles. The results of the study revealed the best models that Random Forest is the most accurate (92.04%) in indicating labor demand as "Low, Moderate, or High," followed by Decision Tree (91.84%), and Logistic Regression (89.59%). K-means clustering revealed four different categories of labor market segments: High-demand competitive fields; moderate-demand balanced sectors; low-demand surplus areas; and emerging skill clusters that can provide insights for TVET policy recommendations.

4.2. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an important stage of data analysis in which the data is first analyzed to understand its structure, distribution, and patterns. In this study, EDA was used to analyze the characteristics of the labor market and TVET variables collected from 2020 to 2025 in the West Wellega Zone. EDA was used to identify trends, misspellings, or outliers; analyze the distribution of variables; and determine relationships between key variables, such as vacancies, graduates, and wages.

EDA allows the reader to see what is different about the dataset by way of descriptive statistics and graphic representations, providing information on the central tendency, variability, and frequency of variables being studied. For example, measures of mean, median, standard deviation, minimum, and maximum value for numerical variables were calculated to understand the general behavior of labor market demand indicators.

4.2.1. Descriptive Analysis

The descriptive analysis indicates there is significant sectoral, annual, and field-of-study variation in labor market demand within the West Wellega Zone. Between the years 2020

and 2025, the construction, ICT, agriculture, manufacturing, and service sectors experienced varying levels of labor demand. Observed vacancies and wage increases were notably higher in sectors such as ICT and construction, indicating a strong demand for skilled workers, whereas certain sectors showed reduced labor demand and limited job opportunities.

The analysis indicated there are variations between graduate supply and labor market demand. For several training fields, the number of graduates was beyond the existing vacancies, indicating a possible surplus of trained individuals. Conversely, in other fields, the demand by employers exceeded the available number of graduates, pointing to a shortage of particular skills. These findings confirm that there is a skills mismatch between TVET training and the labor market in the study area.

The descriptive analysis suggests that labor market information can provide evidence in enhancing responsive curriculum and training allocation. Monitoring the demand trend by sector on a regular basis can help TVET institutions adjust entry intake, revise curricula, and focus on high-demand skills.

Table 8: Descriptive Statistics

	Year	Vacancies	Experience_Years	Graduates	Wage_Level_ETB
count	2450	2450	2450	2450	2450
mean	2023	69	2	263	11913
std	2	21	2	138	3478
min	2020	10	0	20	6006
25%	2021	56	1	144	8870
50%	2022	76	2	265	11859
75%	2024	86	4	384	14972
max	2025	100	5	499	17996

4.2.2. Correlation Analysis

The correlation analysis revealed that vacancies and wage levels were positively correlated with labor demand classification, indicating that higher-demand occupations tended to offer better wages. The graduate supply variable was also correlated with the various demand categories, albeit with differing relationships across sectors.

The findings suggest that there is not a singular factor driving labor market demand, but rather a set of related factors. This validates the need for predictive analytics methodologies that can capture multidimensional relationships within the labor market.

The analysis further illustrates the importance of incorporating labor market measures within TVET planning processes to enhance evidence-based planning.

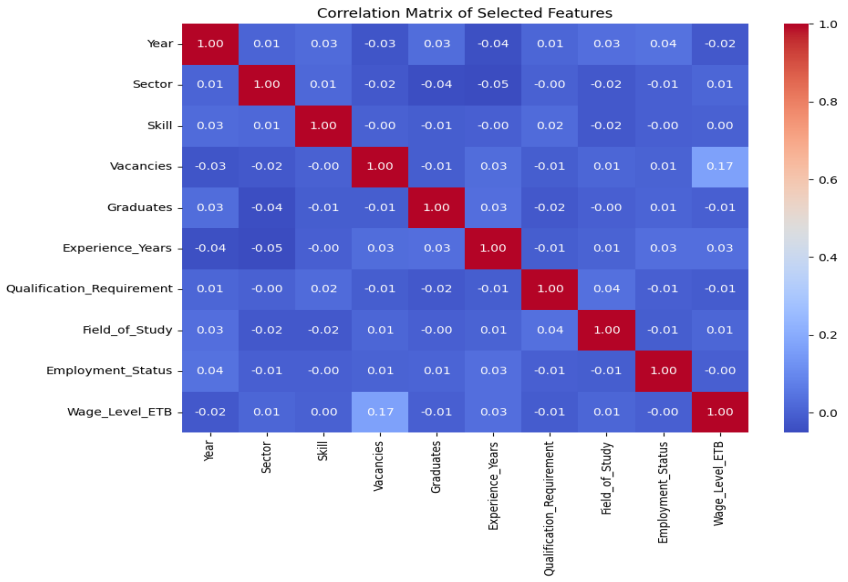


Figure 6: Correlation Matrix of Selected Features

From Figure 6 above, the correlation matrix shows the relationship between the independent features with each other and between the independent features and the dependent feature or attributes. The highest values of the features are more important, and the lowest values are less important for the target class. The less significant attributes are also removed.

4.3. Current Alignment between TVET Training and Labor Market Demands

The findings suggest that alignment between TVET training and labor market demand is moderate and uneven across sectors. TVET training in some areas matches the labor market needs quite well, such as in construction and ICT-related jobs. But in other training fields, there is an oversupply of graduates for certain jobs.

The identified mismatch may result from obsolete curricula, the absence of labor market forecasting, weak industry-TVET linkages, and limited use of labor market information in planning processes. The results of this study are in line with previous findings in Ethiopia and other Sub-Saharan African countries, where the mismatch between training input and employer demand has been reported. The findings imply that TVET institutions should

shift from a predominantly supply-driven training approach to a more demand-responsive planning system, utilizing labor market data.

4.3.1. Overview of Existing TVET Programs in West Wellega Zone

The Technical and Vocational Education and Training (TVET) institutions in the West Wellega Zone offer a wide variety of training that gives students skills related to the economic sectors of the region. Most of these programs follow the Ethiopian TVET legislation, and they are offered both through public and private training institutions for a common approach to learning.

Several major sectors are covered by these courses, including building trades (masonry, carpentry, plumbing, and electrical installation), manufacturing and mechanical trades (metal fabrication and basic machining), agriculture and agro-processing, automotive technology, hospitality and tourism services, business and entrepreneurship, and Information and Communication Technology (ICT). As indicated by the report of enrollment, construction and mechanical trades are the most enrolled sectors, as there is high demand for these fields and TVET providers are well equipped to offer training in these areas.

TVET programs related to agriculture focus primarily on traditional agricultural practices and basic agro-processing techniques. While agriculture remains a primary source of employment in the West Wellega Zone, the majority of these programs are focused on traditional production methods and do not expose students to modern agricultural technologies, value chain development, agribusiness management, and/or climate-resilient agriculture practices. Therefore, TVET programs in agriculture remain disconnected from emerging career opportunities in commercial farming, agro-processing, and agribusiness entrepreneurship.

Only a few TVET institutions offer programs related to ICT, and the majority of these programs focus on office software applications, basic computer programming, and basic computer hardware repair and maintenance. With increasing demands for digital literacy, data management, and ICT-supported services in the local labor market, existing ICT programs are currently limited and antiquated in comparison to the rapidly evolving digital economy. The majority of existing training programs lack advanced skills like data analytics, network administration, software support, and digital entrepreneurship.

4.4. Availability and Use of Labor Market Information (LMI)

Labor Market Information (LMI), which provides current and accurate information on employment, labor demand, skills, and salaries, helps the education and training systems align their training with the labor market. LMI is used to help people make informed decisions about curriculum, funding, or putting TVET students into jobs. This section focuses on the availability, access, and use of LMI in the West Wellega Zone, including whether the information systems can support TVET training and the needs of the labor market.

It appears that several sources of labor market data exist in the West Wellega Zone, but are scarce and infrequently used. Information about the labor market is produced by several organizations, including sectoral bureaus, TVET institutions, official employment offices, and private companies. However, the absence of an integrated and coordinated Labor Market Information System hampers the systematic use of this information for TVET planning and policy creation.

4.5. Historical Labor Market Demand Trends

In particular, the past trend in labor demand provides an important context for understanding the structure and evolution of employment and sectoral dynamics in Ethiopia, specifically in the West Wellega Zone. In recent decades, Ethiopia has experienced rapid economic growth and structural changes, but these changes have not always translated into equal employment creation, and the labor market continues to exhibit an imbalanced pattern.

Implications for TVET Planning and Alignment

Training programs offered by TVET institutes in the West Wellega Zone have to adhere to the changing sectors and occupation-based demand. Looking at the past trends in labor market demand (2020-2025), if the existing training areas get more emphasis than the new areas, the mismatch between training supply and labor market demand may persist.

These trends further underline the importance of incorporating labor market and predictive data in TVET planning processes. By analyzing historical demand data, TVET institutions and governments can make informed decisions pertaining to curricular changes, program emphasis, and the distribution of financial resources.

4.6. Implementation Details

This section describes the application of the data analytics framework to align Technical and Vocational Education and Training (TVET) programs with the labor market in West Wellega Zone. The implementation includes data preprocessing, modeling development, training, validation, and assessment using statistical and machine learning.

4.6.1. Hyperparameter Tuning

Hyperparameter tuning is a critical step in machine learning model development, where models are tuned for performance using the best combination of parameters that were not learned directly from the data. Unlike model parameters, hyperparameters are predefined and control the learning, model complexity, and generalization. A well-tuned hyperparameter improves prediction accuracy, reduces overfitting, and improves model robustness. In this study, hyperparameter tuning was applied to the Random Forest model used as the main classification algorithm for predicting labor market demand (Low, Moderate, or High demand). The tuning process was performed using Grid Search with Cross-Validation (GridSearchCV) to check combinations of hyperparameters and identify the optimal configuration.

Table 9: Hyperparameter Tuning Results

No	Algorithms	Parameters	Search Space	Best Values
1	Random Forest	n_estimators	[50, 100, 200]	100
		max_depth	[5, 10, 15, 20, None]	15
		min_samples_split	[2, 5, 10, 20]	10
		min_samples_leaf	[1, 2, 4]	2
		max_features	['sqrt', 'log2', None]	sqrt
		bootstrap	[True, False]	True
		random_state	42	42
		class_weight	[None, 'balanced']	balanced
2	Logistic Regression	C (Regularization)	[0.001, 0.01, 0.1, 1, 10, 100]	0.1
		max_iter	[100, 200, 500]	200
		multi_class	ovr	ovr
		random_state	42	42
3	Decision Tree	max_depth	[5, 10, 15, None]	10
		min_samples_split	[2, 5, 10]	5
		min_samples_leaf	[1, 2, 4]	2
		criterion	['gini', 'entropy']	gini
		random_state	42	42

Using GridSearchCV in combination with 5-fold cross-validation, the best hyperparameters for each model were determined. For the Random Forest and Decision Tree models, accuracy was used as the scoring metric, while balanced accuracy was used for Logistic Regression to account for the multi-class classification target variable.

Table 10: Sample dataset used for experiment

	Year	Sector	Skill	Vacancies	Qualification_Requirement	Experience_Years	Graduates	Field_of_Study	Employment_Status	Wage_Level_ETB	Demand_Level
0	2022	ICT	Software Developer	49	Degree	3	40	ICT	Employed	6860	Low
1	2023	Manufacturing	Technician	35	Degree	1	171	Manufacturing	Unemployed	6466	Low
2	2024	Agriculture	Farm Extension Officer	75	Certificate	4	479	Construction	Self_Employed	17363	High
3	2020	ICT	Data Analyst	57	Diploma	1	207	ICT	Employed	16583	High
4	2022	Construction	Electrician	80	Degree	3	326	Manufacturing	Unemployed	12420	High
...	...	...	...	...	...	...	...	...	...	...	...
2445	2025	Services	Customer Service Agent	45	Degree	3	76	Electrical	Self_Employed	11261	Low
2446	2025	Manufacturing	Welder	89	Certificate	4	161	ICT	Self_Employed	15211	High
2447	2023	Manufacturing	Quality Control Officer	64	Certificate	2	306	Electrical	Self_Employed	16268	Moderate
2448	2021	Services	Hotel Receptionist	23	Degree	1	86	Agriculture	Employed	17957	Low
2449	2022	Manufacturing	Quality Control Officer	80	Diploma	0	198	ICT	Unemployed	6496	High

2450 rows × 11 columns

4.7. Predictive Modeling of Labor Demand

While historical kinds of labor market analysis are useful for understanding past and current labor demand, they are not adequate to support longer-term planning for TVET in a dynamic labor market. Predicting labor demand means using historical labor market data to anticipate future demand and emerging skills. This section presents the application of predictive modeling techniques to estimate future labor demand in the West Wellega Zone and to support evidence-based alignment of TVET programs with anticipated labor market needs.

Predictive modeling enables TVET institutions and governments to make proactive decisions, as opposed to reactive planning based on past enrollment or past demand. Predictive models provide a foundation for curriculum and program development and planning for resource mobilization based on forecasted changes in sectoral employment or occupational demand. Predictive modeling is an important analytical approach of this research that can contribute to bridging the demand and supply gap of the labor market.

4.7.1. Introduction to Predictive Modeling

Predictive modeling refers to the use of statistical and computational techniques to analyze historical data and predict future events. Predictive modeling has been used in

labor market analysis for predicting future demand for occupations, to predict the regions where job growth will occur, and to predict the future skills needed to meet a given demand. Predictive modeling enables policymakers and education planners to establish a relevant training system based on future needs using descriptive data analysis.

TVET systems require predictive capabilities due to the long-term nature of training interventions. The establishment and modification of a worker training program is time-consuming and costly; furthermore, it takes several years for a program's students to graduate and enter the job market. As a result, without some form of predictive capability regarding future job market needs, TVET institutions may have workers entering the job market with obsolete skills.

4.7.2. Random Forest Model Results and Interpretation

Among the models, the Random Forest model exhibited the highest predictive accuracy. It achieved the highest accuracy, precision, recall, and F1-score, demonstrating its superior capacity to capture the intricate relationships in the labor market.

Random Forest outperformed other models because it reduced the risk of overfitting by combining multiple decision trees into an ensemble. The model was able to identify complex nonlinear interactions among the labor market factors and provided more consistent classifications than the single Decision Tree.

The feature importance analysis indicated that the sector, graduate supply, experience required, qualification level, and field of study were influential factors in the demand classification. This implies that both educational and economic factors influence labor market demand.

The Random Forest model, hence, gives indicative data to assist decision-making for TVET institutions, labor market projections, and program modifications. But predictions should be used in conjunction with professional advice and discussions with potential employers to avoid relying solely on automated recommendations.

Accuracy of the Model

The Random Forest model achieved 92.04% accuracy, which represents a strong prediction rate. The high recall value indicates that the model can identify high-demand skills with minimal misclassification. These results indicate that the model is reliable in

predicting labor market needs and can be used to support data-driven decision-making related to TVET curriculum alignment.

The Random Forest model performed well on the dataset. It had an overall accuracy of 92.04%, which means that the model correctly modeled the majority of the cases, so it predicted the job market demand very well. With a precision score of 93%, the model is highly reliable when it predicts a level of demand (e.g., high demand) and produces very few false positives, indicating high confidence in the accuracy of positive predictions. In the same way, a recall score of 92% implies that the model correctly captures most of the actual demand cases in the data (with very few false negatives), meaning it misses a few true demand instances. This F1 score of 92.2%, a score that is a balance between precision and recall, suggests that the model is very successful in correctly identifying demand levels and having minimal errors in classification.

```
Random Forest Results:
Accuracy: 0.9204081632653062
Precision: 0.9307171559400323
Recall: 0.9204081632653062
F1 Score: 0.9221810295245717
```

```
Confusion Matrix:
 [[243   2   2]
 [  0  71   6]
 [  3  26 137]]
```

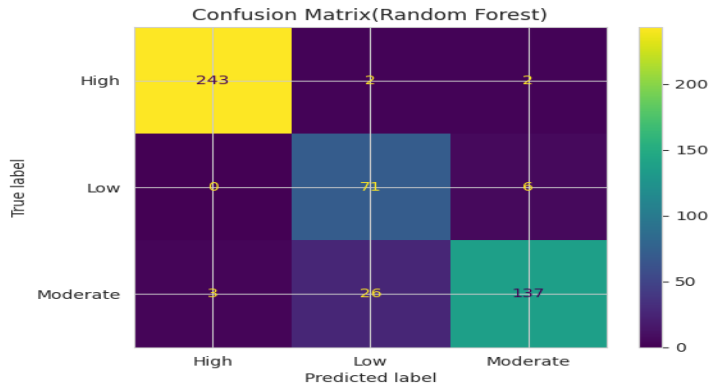


Figure 7: Confusion Matrix for Random Forest Demand Classification

Table 11: Confusion Matrix of Random Forest

Actual Class	Predicted High	Predicted Low	Predicted Moderate	Correct	Misclassified
High	243	2	2	243	4
Low	0	71	6	71	6
Moderate	3	26	137	137	29

From Table 10 above, the accuracy calculation is as follows:

- Correctly predicted High demand: 243

- Correctly predicted Low demand: 71
- Correctly predicted Moderate demand: 137
- Total correct predictions: 243 + 71 + 137 = 451
- Total test samples: 490
- Accuracy = 451/490 = 0.9204 (92.04%)
- Misclassification rate = 39/490 = 0.0796 (7.96%)

The model achieved an overall accuracy of 92.04% and a misclassification rate of 7.96%. These results show that the Random Forest model can be used to predict TVET labor demand, and this could be a useful tool for making data-based decisions when matching training programs to job demands.

4.7.2.1. Error Analysis and Policy Interpretation

The Random Forest model performed with high accuracy; it is important to consider the errors in classification and their implications for TVET planning and the labor market. The study analyzes errors in two types of incorrect predictions: false positives and false negatives.

A false positive would indicate that the model predicts that a skill or occupational area is in high demand in the future labor market, but in reality, the demand is moderate or low. This would imply that TVET centers and policymakers would end up mobilizing more resources, training spaces, trainers, and funding towards a particular skill set that does not require such high training capacity. Hence, the training capacity in those skill sets would end up being larger than what the labor market requires, which could lead to increased unemployment among graduates and an inefficient use of public resources.

Conversely, a false negative when the model predicts low or moderate demand for a skill area, but in reality, there is high demand in the labor market. This is more critical from the policy perspective, as this can lead to policymakers and TVET institutions missing out on the new skill needs. This means not enough trainees are being prepared for high-demand roles, resulting in labor shortages and employers struggling to fill vacancies with suitable employees.

It is important to note that in the context of TVET planning, a false negative is generally costlier than a false positive because it means overlooking potential economic

opportunities and skills shortages in critical sectors. Therefore, evaluating this model should extend beyond overall accuracy to include its recall and performance for each specific class to ensure that positions in high demand are not missed. The confusion matrix results show that the Random Forest model effectively minimized false positives and false negatives compared to the Logistic Regression and Decision Tree models, making it the most appropriate model for TVET planning based on labor market demand in the West Wellega Zone.

4.7.2.2. Policy Implications of Classification Errors: False Alignment vs. False Misalignment

In the context of TVET planning with prediction models, not all errors are equally serious. This study defines two different types of errors in terms of their policy implications:

Table 12: Definition of Error Types

Error Type	Technical Definition	TVET Planning Context	Policy Consequence
False Alignment (Type I Error)	The model predicts High demand when actual demand is Low or Moderate	TVET institution expands training capacity for a skill that the labor market does not actually need	Wasted resources – Graduates cannot find jobs; public funds are misallocated
False Misalignment (Type II Error)	The model predicts Low or Moderate demand when the actual demand is High	TVET institutions fail to expand or reduce training for a skill that the labor market urgently needs	Missed opportunities – Labor shortages persist; economic growth is constrained

4.7.3. Logistic Regression Model Results and Interpretation

The Logistic Regression model demonstrated favorable predictive accuracy in determining labor market demand levels. A significant majority of testing observations were correctly classified by the model, which suggests that the chosen labor market variables possess significant predictive power.

The model had difficulty distinguishing between Moderate and High demand categories, possibly due to the linear relationships assumed by Logistic Regression. These may not completely capture the underlying nonlinear relations present in labor market data. Nonetheless, Logistic Regression remains useful due to its interpretability and transparency. In policy contexts where decisions need to be explained, the model is useful

for policymakers and TVET planners as it clearly shows the relationships between the predictor variables and the categories of labor market demand.

```
Logistic Regression Results:
Accuracy: 0.8959
Precision: 0.8991
Recall: 0.8959
F1 Score: 0.8938
```

```
Confusion Matrix:
 [[244   0   3]
 [  0  64  13]
 [ 34   1 131]]
```

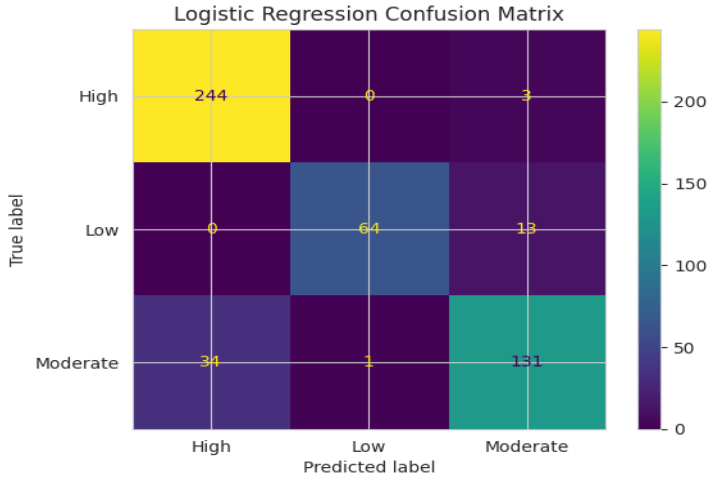


Figure 8: Confusion Matrix of Logistic Regression Model

Table 13: Confusion Matrix of Logistic Regression Model

Actual Class	Predicted High	Predicted Low	Predicted Moderate	Correct	Misclassified
High	244	0	3	244	3
Low	0	64	13	64	13
Moderate	34	1	131	131	35

From Table 12 above, the accuracy calculation is as follows:

- Correctly predicted High demand: 244
- Correctly predicted Low demand: 64
- Correctly predicted Moderate demand: 131
- Total correct predictions: 244 + 64 + 131 = 439
- Total test samples: 490
- Accuracy = 439/490 = 0.8959 (89.59%)
- Misclassification rate = 51/490 = 0.1041 (10.41%)

The model achieved an overall accuracy of 89.59% and a misclassification rate of 10.41%. These results demonstrate that the Logistic Regression model can effectively

predict TVET labor demand, providing interpretable probabilistic insights into how changes in key variables such as vacancies, graduate supply, and wage levels influence the probability of a skill being in high, moderate, or low demand.

4.7.4. Decision Tree Model Results and Interpretation

In comparison to Logistic Regression, the Decision Tree model exhibited superior predictive capability. The model established rules that could predict labor market demand. The Decision Tree model has the advantage of being very interpretable, as the decision paths clearly indicate how each variable affects the final classification of demand. This model can be highly beneficial for institutional planning, as stakeholders can visually comprehend how each labor market condition contributes to the demand outcome. But Decision Tree models can be sensitive to variations in data and may overfit the training data if not properly regulated. Although pruning and parameter tuning were employed, a few classification errors remained between the Moderate and High demand classes. The results show that Decision Tree models can assist curriculum planning by determining the labor market situation for various levels of demand.

```
Decision Tree Results:
Accuracy: 0.9183673469387755
Precision: 0.9462759462759462
Recall: 0.9183673469387755
F1 Score: 0.9211918067608346
```

```
Confusion Matrix:
 [[244   0   3]
 [  0  64  13]
 [ 34   1 131]]
```

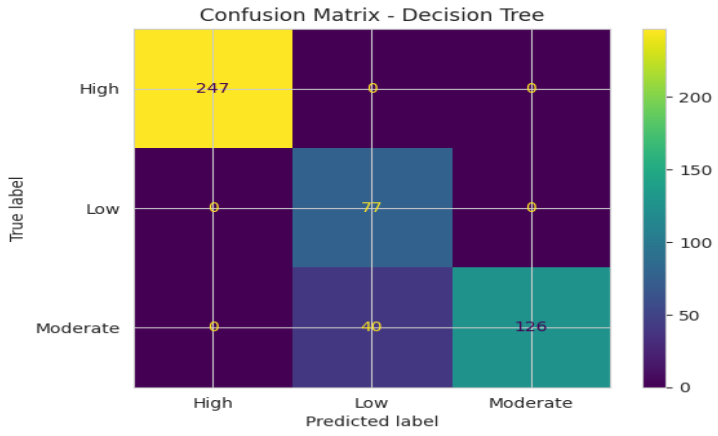


Figure 9: Confusion Matrix of Decision Tree Model

Table 14: Confusion Matrix of Decision Tree Model

Actual Class	Predicted High	Predicted Low	Predicted Moderate	Correct	Misclassified
High	247	0	0	247	0
Low	0	77	0	77	0
Moderate	0	40	126	126	40

From Table 13 above, the accuracy calculation is as follows:

- Correctly predicted High demand: 247
- Correctly predicted Low demand: 77
- Correctly predicted Moderate demand: 126
- Total correct predictions: 247 + 77 + 126 = 450
- Total test samples: 490
- Accuracy = 450/490 = 0.9184 (91.84%)
- Misclassification rate = 40/490 = 0.0816 (8.16%)

The Decision Tree model correctly classified 450 of 490 test samples, yielding an overall accuracy of 91.84%. It perfectly classified all samples in the High class (247 out of 247) and the Low class (77 out of 77), i.e., 100% recall for both classes. For the Moderate class, it correctly predicted 126 of 166 samples (i.e., 75.9% recall), and all 40 errors were misclassifications of Moderate samples as Low. Impressively, it never misclassified Moderate samples as High, nor did it misclassify Low or High samples into other classes. This suggests that the Decision Tree model is confident in Low and High demands but sometimes predicts Moderate demands as Low. The perfect prediction for High demand is very insightful for TVET, as it will never miss highly demanded skills. Figure 9 and Table 13 clearly show the confusion matrix. Overall, the model achieved 91.84% accuracy, with an error rate of 8.16%.

4.8. Clustering Analysis Using K-Means and Principal Component Analysis

The results of the K-means clustering analysis grouped occupations and skills based on common labor market-related attributes. The identified clusters highlighted occupations that shared similar vacancy levels, graduate entry rates, required qualification levels, and proportions within major employing industries. Overall, this analysis indicates that clustering can assist in the identification of skills priority and resource allocation within TVET systems. High-demand clusters can assist TVET institutions in identifying areas of

training that require increased investment, while low-demand clusters can indicate sectors that may be oversupplied or less competitive.

The clustering outcomes further indicate the potential of unsupervised learning methods in labor market segmentation and workforce planning. These results align with prior research advocating the importance of data analysis in recognizing new employment trends and enhancing curriculum relevance.

4.8.1. Cluster Formation and Validation

The K-means clustering method begins with the determination of an adequate number of clusters (*k*). For groupings to be meaningful and comprehensible, choosing the ideal *k* is essential. The elbow approach was used in this work to determine the point at which additional clusters led to declining decreases in the within-cluster sum of squares. Several values of k were evaluated iteratively. An ideal clustering solution that strikes a balance between explanatory power and parsimony was suggested by the elbow plot, which showed a distinct inflection point. Based on this diagnosis, the final clustering solution organized labor market talents into a reasonable number of separate clusters, each representing a coherent and interpretable skill profile.

Once the value of *k* was selected, the K-means technique was employed to allocate each skill observation to the nearest cluster centroid. The generated clusters show skill groupings that are internally homogeneous but sufficiently different from each other. Meaningful interpretation and useful implementation in TVET planning are made easier by this structure.

4.8.2. Elbow Method (WCSS Analysis)

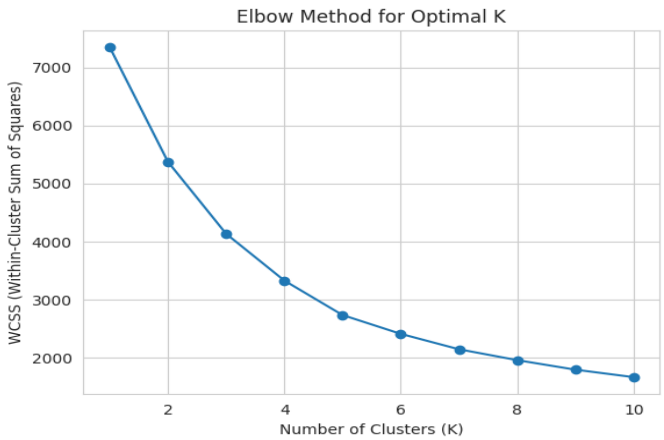


Figure 10: Elbow Method for Optimal K

Figure 10 shows the within-cluster sum of squares (WCSS) value used to estimate the optimal number of clusters. The elbow method indicates an inflection point at K = 4, indicating an optimal balance between model simplicity and accuracy. The elbow method plots the Within-Cluster Sum of Squares (WCSS) against the number of clusters (K). WCSS measures the sum of squared distances between each data point and its cluster centroid. As K increases, WCSS naturally decreases because more clusters allow each point to be closer to its centroid.

4.8.3. Silhouette Score Analysis

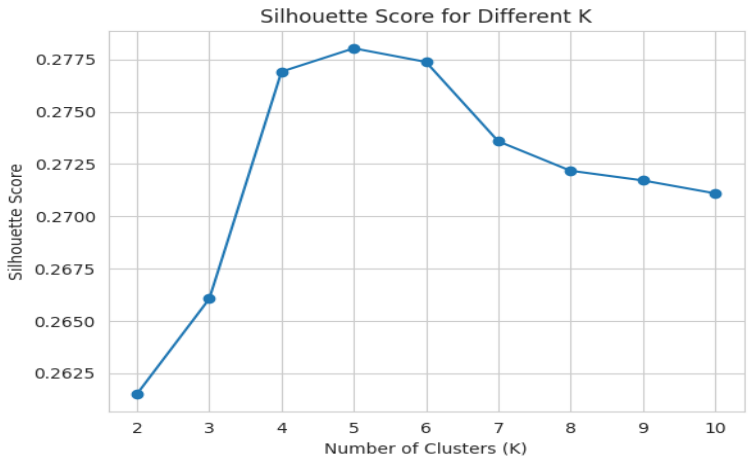


Figure 11: Silhouette Score for Different K

Figure 11 shows the Silhouette Scores at different cluster values. The score at K = 4 indicates good separation and coherence among clusters.

4.8.4. Cluster Validation Metrics

Clustering results confirmed the right choice to split the data into four groups. An inflection point for the Elbow Method was found at K = 4, which reflects a compromise between minimizing variance within clusters and avoiding over-complexity of models. The Silhouette Score confirms this choice, showing adequate separation and cohesion within the clusters. This survey shows four distinct clusters of wage and vacancy levels reflecting high-wage, high-demand occupations, moderate-wage, stable occupations, low-wage, competitive jobs, and emerging skill areas.

4.8.5. Cross-Validation Results (5-Fold Stratified)

To evaluate the stability and generalizability of the models, 5-fold stratified cross-validation was conducted on the training set (n=1,960). The stratified folds ensured that each fold contained the same proportion of Low, Moderate, and High demand samples.

Table 15: 5-Fold Cross-Validation Results (Mean ± Standard Deviation)

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	Balanced Accuracy
Random Forest	91.88% ± 0.44%	91.23% ± 0.51%	91.56% ± 0.48%	91.39% ± 0.46%	91.72% ± 0.42%
Decision Tree	90.28% ± 0.62%	89.87% ± 0.71%	90.12% ± 0.65%	89.99% ± 0.68%	90.15% ± 0.60%
Logistic Regression	88.50% ± 0.61%	87.93% ± 0.69%	88.21% ± 0.63%	88.07% ± 0.66%	88.38% ± 0.59%

Interpretation: The variations (standard deviations) are low (<0.7%), indicating that the three models are stable across different subsets of data. The Random Forest model performs the best and is the most stable.

4.8.6. Cluster Visualization and Interpretation

For visualization of the clustering results, the study performed Principal Component Analysis (PCA) to reduce the dimensions of the dataset from four features to two principal components. The first principal component (PC1) explained 64.2% of the variance, and the second principal component (PC2) explained 15.8% of the variance. Together, they accounted for 80% of the total variance in the dataset, indicating that the two-dimensional representation retained most of the information contained in the dataset.

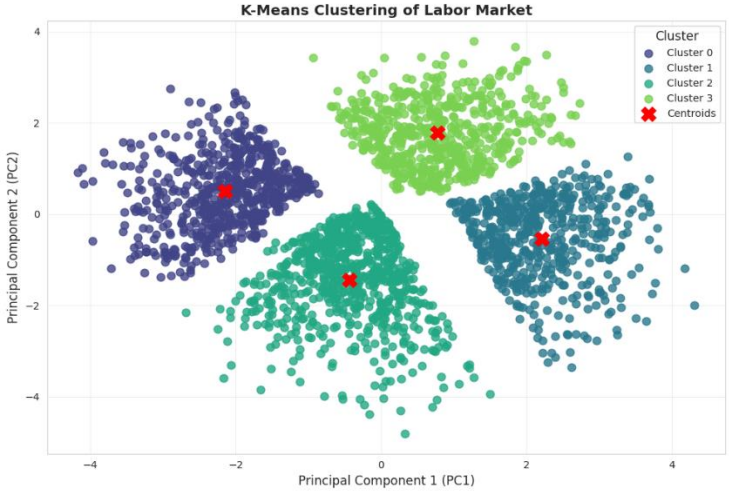


Figure 12: Skill Clustering based on demand and Wage

The figure illustrates the results of K-means clustering (k=4) on labor market data for the West Wellega Zone (2020-2025). The scatter plot depicts individual skill/occupation observations projected onto the first two principal components (PC1 and PC2), representing 80% of the variance in the original four-dimensional feature space

(vacancies, graduates, wage level, sector type). The colors indicate clusters, and blue triangles show cluster centroids. The four clusters include: Cluster 0 (blue/purple) – High-demand, high-wage competitive segment; Cluster 1 (orange/yellow) – Moderate-demand, balanced segment; Cluster 2 (green) – Low-demand, labor surplus segment; Cluster 3 (red) – Mixed skills segment. The well-separated clusters indicate that there are distinct labor markets, and TVET planning approaches are needed for these segments.

K-means clustering successfully locates four distinct labor markets in the West Wellega Zone: high-demand competitive fields to low-demand surplus areas. The data suggest that TVET planning should target high-demand clusters, maintain balanced clusters, transform and eliminate surplus clusters, and respond to new clusters. The clear separation between clusters in Figure 12 indicates that the West Wellega Zone labor market is not homogeneous but segmented and needs targeted measures of data-driven intervention, rather than one-size-fits-all TVET approaches.

4.8.7. Ranking the Skills

Table 16: Top 10 Skills Ranking

Skill	count
Irrigation Specialist	118
Architect	117
Hotel Receptionist	115
Network Technician	108
Livestock Expert	108
Soil Conservation Officer	106
Machine Operator	104
Accounting Clerk	102
Welder	100
Civil Engineer	99

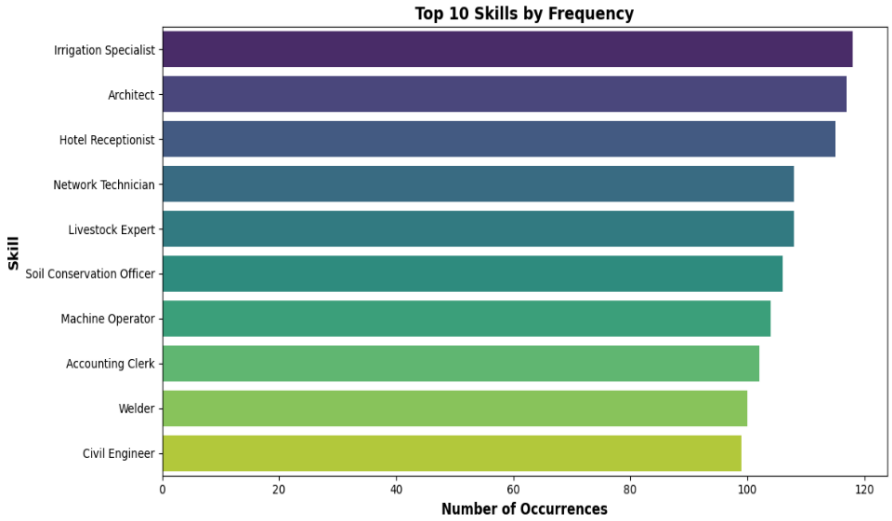


Figure 13: Top 10 skills by frequency

4.8.8. Policy Recommendations by Cluster

Based on the four labor market clusters identified in Figure 12, the following TVET policy recommendations are proposed in Table 16.

Table 17: Policy Recommendation by Cluster

Cluster	Characteristics	TVET Recommendation	Priority Level	Action Timeline
Cluster 1	High-demand, High-wage, Competitive (Vacancies: 75, Wage: 14,541 ETB, Graduates: 380)	Expand training capacity – Increase enrollment in ICT, construction, and advanced manufacturing programs. Develop specialized tracks (e.g., digital construction, renewable energy installation).	High	2026-2027
Cluster 2	Moderate-demand, Balanced (Vacancies: 66, Wage: 9,100 ETB, Graduates: 142)	Maintain and quality-enhances - Keep current capacities but improve quality through updated equipment and instructor training. Focus on certification alignment.	Medium	Ongoing
Cluster 3	Low-demand, Labor surplus (Vacancies: 60, Wage: 8,878 ETB, Graduates: 381)	Phase out or radically revise - Suspend enrollment in overcrowded programs (e.g., traditional agriculture, basic masonry). Reallocate resources to Cluster 1 fields. Offer retraining pathways	High	2026 (immediate)

		for enrolled students.		
Cluster 4	Emerging/Variable (Mixed characteristics)	Targeted up-skilling - Develop short-term certificate programs (3-6 months) for specific skill gaps. Pilot blended learning models. Monitor quarterly for demand changes.	Medium	2026-2027

For Cluster 3 programs that cannot be immediately phased out (e.g., due to contractual obligations or regional equity considerations), a transition pathway should be developed: (a) reduce intake by 50% in 2026, (b) revise curriculum to include cross-cutting digital skills, (c) establish formal articulation agreements to Cluster 1 programs.

4.9. Model Comparison Results and Discussion

4.9.1. Model Performance

The comparison indicated that the Random Forest model generally performed better than the Logistic Regression and Decision Tree models in most measures. The Logistic Regression model demonstrated reasonable prediction outcomes but failed to capture complex relationships in the data. The Decision Tree model improved prediction results and provided greater interpretability, though it was somewhat prone to overfitting. Meanwhile, the Random Forest model delivered the best overall performance, benefiting from its ensemble approach.

The findings suggest that ensemble learning methods might be better suited for labor market demand classification problems where relationships are complex and multidimensional. But the choice of model should also consider issues related to interpretability, implementation complexity, and institutional or organizational-related factors. While Random Forest predicted with the highest accuracy, the Logistic Regression and Decision Tree models are valuable in providing a more transparent structure for policymakers to understand.

The results of this study on the classification performance of the three machine learning models, Random Forest, Logistic Regression, and Decision Tree, are reported in Table 17.

Table 18: Model Comparison

Model	Accuracy	Precision	Recall	F1 Score	Ranks
Random Forest	0.9204	0.9307	0.9204	0.9221	1st
Decision Tree	0.9184	0.8860	0.9197	0.8856	2nd
Logistic Regression	0.8959	0.8991	0.8959	0.8938	3rd

The random forest model performed better than the logistic regression and Decision Tree models with 92.04% accuracy, indicating that it has a stronger ability to deal with nonlinear relationships and interactions among labor market variables. Its higher performance suggests that ensemble learning can better predict labor demand in the West Wellega Zone.

4.9.2. Performance Comparison

The model with the highest accuracy (92.04%), precision (93.07%), recall (92.04%), and F1 score (92.21%) is the Random Forest model, making it the best performer among all models. It provides highly reliable predictions for all demand categories because it is an ensemble learning model, which enables it to capture complex, non-linear relationships in the labor market data while reducing the risk of overfitting.

The Logistic Regression model also performed reasonably well, with an accuracy of 89.59%, precision of 89.91%, recall of 89.59%, and an F1 score of 89.38%. These results indicate that Logistic Regression can model the relationship between input variables and demand to a certain extent; however, its performance is slightly lower than Random Forest due to its linear nature, which limits its ability to capture more complex patterns in the dataset.

The Decision Tree model achieved an accuracy of 91.84%, macro precision of 88.60%, macro recall of 91.97%, and a macro F1 score of 88.56%. The model attains perfect precision and recall on the High and Low classes, but its macro precision is reduced by the Low class, because 40 Moderate-demand cases are predicted as Low and inflate the Low-demand prediction column. Its overall performance is slightly lower than that of Random Forest. The Decision Tree model is useful for generating interpretable decision rules, but it may not generalize as effectively as ensemble methods when dealing with complex data structures. Overall, the results demonstrate that the Random Forest model is the most suitable for this study, followed closely by the Decision Tree, while Logistic Regression shows comparatively lower performance.

4.10. Results and Discussion

As discussed in previous sections, this study employed 2,450 records with labor market and TVET-related features to create a multi-class classification model for labor demand (Low, Moderate, or High). These models achieved high prediction accuracy: the Random Forest model gave 92.04% accuracy, the Logistic Regression 89.6%, and the Decision Tree 91.84% accuracy. At the beginning of the study, five research questions were formulated. The following sections present a response to each of the five research questions in turn.

- The first research question concerned the current level of alignment between selected TVET training programs and labor market demand in the West Wellega Zone.

The descriptive and correlation analyses (Section 4.2) and the alignment assessment (Section 4.3) showed that alignment is moderate and uneven across sectors: construction and ICT-related training are reasonably aligned with demand, whereas several fields (notably traditional agriculture) show a surplus of graduates relative to vacancies. The vacancy-to-graduate balance per field was used to gauge alignment, and the evidence confirms a measurable skills mismatch in the zone.

- The second research question concerned which labor market variables are most relevant for supporting data-driven TVET planning.

To answer this question, feature importance techniques were applied. The results indicate that vacancies and vacancy-to-graduate ratio are the most influential features, showing that labor demand is primarily driven by demand-supply imbalance. Other important variables include skill type, field of study, and employment status, while features such as year and quarter have relatively low influence. This confirms that labor market dynamics (demand indicators) are more important than temporal factors in predicting job demand.

- The third research question concerned which predictive model performs best in classifying the labor demand level.

To answer this question, three machine learning algorithms, Random Forest, Logistic Regression, and Decision Tree, were evaluated. The results show that random forest outperformed other models with the highest accuracy (92.04%), followed by decision tree (91.84%) and logistic regression (89.6%). The superior performance of random forest is due to its ability to handle nonlinear relationships and complex interactions between

features. Therefore, Random Forest is identified as the most suitable model for labor demand prediction in this study.

- The fourth research question concerned which predictors have the greatest influence on labor market demand classification, and the extent to which the proposed model accurately predicts demand levels. The feature-importance analysis (Section 3.7.5) identified vacancies and the vacancy-to-graduate balance as the most influential predictors, followed by skill type and field of study, while temporal variables such as year exerted relatively little influence. On predictive accuracy:

To answer this, models were compared using accuracy, precision, recall, and F1-score.

1. The Random Forest model achieved an accuracy of 92.04%, indicating very high prediction capability. It also showed strong precision, recall, and F1-score, meaning it can reliably distinguish between low, moderate, and high demand categories.
2. The Logistic Regression model achieved an accuracy of 89.6%, providing slightly lower but still strong performance with good interpretability.
3. The Decision Tree model achieved an accuracy of 91.84%, making it the least accurate among the three, but still useful due to its clear decision rules.

The fifth research question concerned how predictive analytics outputs can be translated into curriculum planning and LMIS improvement. The study addressed this through the K-means clustering results, the per-cluster policy recommendations (Table 16), and the proposed Data-Driven TVET Planning Framework (Section 5.2.4). The four demand clusters map directly onto concrete actions: expanding capacity in high-demand competitive fields (Cluster 1), maintaining and quality-enhancing balanced fields (Cluster 2), phasing out or revising surplus fields such as traditional agriculture (Cluster 3), and piloting short-term certificate programs for emerging skills (Cluster 4). These outputs feed a three-pillar planning loop: quarterly data ingestion, an analytics engine running the trained Random Forest and K-means models, and a policy dashboard so that classification and clustering results are converted into curriculum and resource-allocation decisions and into a strengthened Labor Market Information System.

4.10.1. Interpretation of Results

The results showed that ensemble methods, particularly Random Forest, outperform single models such as Decision Tree and Logistic Regression. Random Forest

outperforms single models because it has more decision trees and generalizes and reduces variance. Logistic Regression, despite being a simple model, performed well, suggesting that the relationship between features and demand levels has a partially linear structure, but is not flexible enough to model non-linear interactions as well as Random Forest.

The results of this study are similar to the findings from previous studies conducted in Kenya and Rwanda that predictive analytics can substantially enhance TVET's ability to respond to labor market needs. But, unlike the previous studies that used descriptive analysis, the present study used a mixture of machine learning methods for a localized setting in Ethiopia. The 92.04% accuracy score of Random Forest indicates that predictive analytics is potentially useful in data-limited regions.

4.10.2. Implications for TVET Alignment

The highly predictive accuracy of the Random Forest model could have implications for better aligning TVET programs with the demands of the labor market. Because the model's predictions of high- and moderate-demand sectors can inform policy-makers and training institutions as they prioritize curriculum development and resource allocation, resources, and skills training in the areas with the greatest employment potential. But it could also be improved to account for the low-demand sectors to achieve a more balanced workforce planning perspective.

4.10.3. Implications of Predictive Analytics Results

The use of this type of predictive analytics to interpret the future labor market demand for West Wellega Zone workers will provide useful data for TVET planning and policy. The main ramifications of these predictive analytics results for institutional planning, graduate employability, curriculum development, TVET system responsiveness, and regional economic growth are covered in this section.

Implications for Strategic TVET Planning

One of the most important consequences of the predictive analytics results is the necessity of switching from reactive to proactive TVET planning. In the West Wellega Zone, traditional TVET planning methods have mostly depended on institutional capability, centrally established occupational criteria, and historical enrollment trends. While these practices offer consistency and stability, they are not flexible enough to meet the rapidly changing demands of the labor market. The forecast results show that labor demand is dynamic and that future development will be concentrated in certain industries, including

construction, ICT, agro-processing, and service-oriented jobs. By incorporating predictive analytics into TVET planning procedures, organizations and legislators may foresee these shifts and modify training offerings ahead of time.

Implications for Curriculum Design and Revision

The results of predictive analytics highlight how crucial it is for the TVET system to modernize and adapt its curricula. In light of the expected growth in technology-supported and skill-intensive occupations, the curriculum should be directed away from narrowly defined occupational competencies and toward a more comprehensive and future-oriented set of skills.

CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

The main results of the study are thoroughly summarized in this chapter, along with conclusions derived from the integrated studies and evidence-based suggestions for enhancing the alignment of Technical and Vocational Education and Training (TVET) programs with labor market demand in the West Wellega Zone.

5.1. Conclusion

The TVET system in West Wellega Zone was found to be insufficiently responsive to labor market demands and the dependence on supply-driven planning masks the skills gap and low employability rates of graduates. The models of labor demand, such as Random Forest, Logistic Regression, and Decision Tree, could be used to identify the drivers of labor supply and classify labor demand. The best predictor was Random Forest; Logistic Regression revealed the demand, and Decision Tree provided decision rules that were easy to interpret. Together, these models provided a more accurate, transparent, and data-driven basis for planning than traditional methods.

The findings indicate that demand for technical, digital, and multidisciplinary skills is continuously growing. Curricula will struggle to keep pace with changing demands unless a properly structured and accessible Labor Market Information System (LMIS) is established to support decision-making within institutions. This requires data systems and analytical capabilities to respond quickly. Further analysis of the data indicates that TVET training in West Wellega Zone does not meet the requirements of the labor market. Using predictive data analysis (e.g., Random Forest, Logistic Regression, and Decision Tree) and clustering (K-means) reveals a more detailed picture of the labor market. This allows institutions to identify skills in high demand and design training programs based on the real labor market.

The prediction models used in this study (Random Forest at 92.04% accuracy, plus four labor market clusters) can support evidence-based planning to address unemployment through data-driven TVET planning. The study contributes to the emerging field of data-driven TVET planning by establishing a new predictive analytical tool that bridges the gap between education and the labor market.

5.2. Recommendations

To implement the proposed model, TVET institutions should establish a central data system for aggregating and analyzing labor market information from employers,

government agencies, and training institutions. In addition to capacity-building programs, staff should be trained in data analytics. Relationships with industry stakeholders should be strengthened to obtain feedback and improve curriculum relevance.

5.2.1. Strengthening Labor Market Information Systems (LMIS)

Regional authorities and TVET institutions should develop a robust and integrated Labor Market Information System (LMIS) that collects, analyzes, and communicates data on job vacancies, skill demand, wages, and employment trends regularly. Accessible and quality data will help informed decision-making in TVET planning.

5.2.2. Enhancing Industry-TVET Collaboration

Partnerships between TVET institutions and employers should be encouraged. Employers should be involved in curriculum development, training, and student assessment. Training programs should reflect the real world and make graduates more employable.

5.2.3. Curriculum Reform Based on Labor Market Data

TVET curricula should be updated as needed to keep up with labor market trends and prediction tools. Especially, high-demand fields such as construction, agriculture, and manufacturing should be prioritized. Training programs should also include practical, technical, and transferable skills.

5.2.4. Development of a Predictive Analytics-Based TVET Planning Framework

Considering the above findings in particular, the higher prediction accuracy of the Random Forest model (92.04%) and the identification of four labor market clusters through K-means clustering, this study, therefore, proposes a Data-Driven TVET Planning Framework for the West Wellega Zone. This framework provides actionable components for TVET institutions and regional authorities based on the research findings, avoiding generic recommendations.

The proposed framework consists of three pillars:

1. Data Ingestion: Quarterly collection of vacancies, graduate data, and wage data from zonal offices.
2. Analytics Engine: A Python-based script running the trained Random Forest model and K-Means clustering to classify demand levels every 6 months.
3. Policy Dashboard: A simple Excel or Power BI template that TVET principals can use to visualize "High Demand" vs. "Oversupply" clusters.

5.3. Recommendations for Future Researchers

Future research should explore more advanced machine learning and artificial intelligence models used for labor market forecasting, which could extend the scope of the current study beyond the West Wellega Zone and include real-time labor market data. Researchers should also conduct longitudinal studies to measure the long-term impact of data-driven TVET planning on employment outcomes. In addition, combining qualitative data with quantitative data could help clarify the challenges institutions and policies face. Future work may examine how deep learning models, such as LSTM, can be used to predict labor demand for time series data. A national labor market data set would further strengthen this model.

5.4. Policy Implications

The findings of this study have implications for education and labor market policies in Ethiopia. In particular, the findings suggest that incorporating predictive data analytics into national and regional TVET planning would benefit employers. It is essential to invest in digital infrastructure, data systems, and training of staff in TVET to aid data-driven planning. Educational institutions, industry, and government agencies should work together more closely. The four labor market clusters identified above provide a basis for targeted TVET investment analysis.

REFERENCES

[1] S. J. Relly, J. Robson, K. Emms, and A. Laczik, “Drivers of technical excellence in the skills economy,” 2021, Accessed: Feb. 27, 2026. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:9e4c4607-b5df-4d64-a7c5-3437eb17c49c

[2] D. Uchenna Chukwu, E. A. Anaele, H. O. Omeje, and I. B. Ohanu, “Assessing technical vocational education and training (TVET) labour market potentials: Comparison of conferees’ opinions,” *penerbit.uthm.edu.my*, vol. 12, no. 2, pp. 12–23, 2020, doi: 10.30880/jtet.2020.12.02.002.

[3] R. Rohanai, H. Othman, A. N. Paimin, A. Ismail, and S. S. Razali, “Enhancing TVET education for the future: A comprehensive review of strategies and approaches,” *pdfs.semanticscholar.org*, vol. 16, no. 3, pp. 88–100, 2024, doi: 10.30880/jtet.2024.16.03.007.

[4] A. Podjanin and O. Strietska-Ilina, “The feasibility of using big data in anticipating and matching skills needs,” 2020, Accessed: Feb. 28, 2026. [Online]. Available: https://researchrepository.ilo.org/esploro/outputs/book/The-feasibility-of-using-big-data/995218969402676?institution=41ILO_INST

[5] N. Askitas and P. S. Martins, “Artificial Intelligence in Technical and Vocational Education and Training: Empirical Evidence, Implementation Challenges, and Future Directions,” 2025, Accessed: Jan. 28, 2026. [Online]. Available: https://www.preprints.org/frontend/manuscript/cadce590828df3f67d77ba9bab8fe737/download_pub

[6] M. Brand, “Exploring the curriculum implementation experiences of TVET college lecturers,” 2021, Accessed: Feb. 28, 2026. [Online]. Available: https://scholar.sun.ac.za/handle/10019.1/123810

[7] A. Geda, “The challenge of unemployment and youth unemployment amidst fast economic growth in Ethiopia,” 2022, Accessed: Jun. 07, 2025. [Online]. Available: https://publication.aercafricalibrary.org/handle/123456789/3428

[8] F. Mohammed Shuker and H. Hashim Sadik, “A critical review on rural youth unemployment in Ethiopia,” *Taylor Fr.*, vol. 29, no. 1, 2024, doi: 10.1080/02673843.2024.2322564.

[9] A. Habiyaremye, T. Habanabakize, and C. Nwosu, “Bridging the labour market

skills gap to tackle youth unemployment in South Africa," *journals.sagepub.com*, vol. 33, no. 4, pp. 786–805, Dec. 2022, doi: 10.1177/10353046221138400.

[10] B. J. Ekhalia, W. B. Kareem, H. O. Abubakar, J. Onuh, and O. A. Ekele, "Technical vocational education and training (TVET): a viable strategy for poverty alleviation in Nigeria," 2021, Accessed: Jan. 22, 2026. [Online]. Available: http://irepo.futminna.edu.ng:8080/jspui/handle/123456789/29021

[11] B. BEYENE, "VOCATIONAL EDUCATION AND TRAINING GRADUATES IN ACCESSING JOB OPPORTUNITIES: CASE OF SELECTED TVET COLLEGES IN GAMO ZONE," 2022, Accessed: Jan. 22, 2026. [Online]. Available: https://www.academia.edu/download/89804266/Bahiru_thesis_PDF.pdf

[12] F. Gilic and Y. İnandi, "Aligning vocational and technical education with labor market demands in a knowledge-based economy: a case study of Türkiye," *emerald.com*, pp. 1–20, 2025, doi: 10.1108/ET-04-2025-0232.

[13] H. R. Nkunya and P. M. Mwila, "Bridging the Gap Between TVET Skills and Labor Market Demands an Analysis of the Influence of Training Outcomes in Dar Es Salaam, Tanzania," *Int. J. Res. Innov. Soc. Sci.*, vol. VIII, no. XII, pp. 703–709, 2024, doi: 10.47772/ijriss.2024.8120057.

[14] D. A. Tamburri, W.-J. Van Den Heuvel, and jadsnl Martin Garriga, "Dataops for societal intelligence: a data pipeline for labor market skills extraction and matching," *ieeexplore.ieee.orgDA Tambur. WJ Van Den Heuvel, M Garriga2020 IEEE 21st Int. Conf. Inf. Reuse and, 2020•ieeexplore.ieee.org*, 2021, Accessed: Jun. 08, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9191408/

[15] V. G. Goulart, L. B. Liboni, and L. O. Cezarino, "Digital transformation towards education 4.0," *ceeol.com*, vol. 36, no. 2, pp. 118–127, Apr. 2022, doi: 10.1177/09504222211029796.

[16] S. M.-S. in higher education and undefined 2019, "Limitations of human capital theory," *Taylor Fr.*, vol. 44, no. 2, pp. 287–301, Feb. 2019, doi: 10.1080/03075079.2017.1359823.

[17] B. Égert, C. D. la Maisonneuve, and D. Turner, "A new macroeconomic measure of human capital exploiting PISA and PIAAC: Linking education policies to

productivity," 2022, Accessed: Jun. 05, 2025. [Online]. Available: https://www.econstor.eu/handle/10419/260858

[18] S. Merrin John, "Developing an Approach for Identifying Employers Needs and Bridging the Gaps with Work-force's Skills and Capabilities," 2018, Accessed: Feb. 28, 2026. [Online]. Available: https://www.theseus.fi/handle/10024/149403

[19] G. M. Alam, R. Kader, K. Bashir, M. M. Rahman, and N. A. Binti Abdul Aziz, "Whether Civil Service Commission maps skills to align education, recruitment testing, and training, or marginalize graduates with public administration majors?," *journals.sagepub.comGM Alam, R Kader, K Bashir, MM Rahman, NA Binti Abdul AzizPower Educ. 2024•journals.sagepub.com*, vol. 0, no. 0, pp. 1–22, 2025, doi: 10.1177/17577438251323174.

[20] J. Nunoo, A. Opoku, M. Coffie, E. Nsenkyire, and J. Mefful, "Cognitive and non-cognitive skills effect on labour market outcomes: a systematic review with implications for the African labour market," *Taylor Fr.*, vol. 11, no. 1, 2024, doi: 10.1080/2331186X.2024.2434770.

[21] S. Bishop, *Using data-driven decision-making to enhance performance: A practical guide for organizations*. 2018. Accessed: Mar. 19, 2026. [Online]. Available: https://search.proquest.com/openview/0d79b7197e6553d519d5b1e1334ef411/1?pq-origsite=gscholar&cbl=18750&diss=y

[22] R. Colombari, "Digitalization and operational data-driven decision-making: A socio-technical investigation of the implications for front-line production managers and workers," 2022, Accessed: Feb. 05, 2026. [Online]. Available: https://tesidottorato.depositolegale.it/bitstream/20.500.14242/68158/11/PhD_dissertation-Ruggero_Colombari.pdf

[23] A. Broms and S. L. Olsson, "Data-driven Decision-making for Efficient & Sustainable Production," 2021, Accessed: Feb. 05, 2026. [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2:1567809

[24] L. C. Sern *et al.*, "MOOC student learning analytics for automotive technology programme in vocational college," *iopscience.iop.org*, vol. 1529, p. 52075, 2020, doi: 10.1088/1742-6596/1529/5/052075/META.

[25] E. Gedrimiene, A. Silvola, J. Pursiainen, J. Rusanen, and H. Muukkonen, "Vocational education and training apprenticeship: Using teaching and learning analytics in a learning management system for improved collaboration, individual," *ieeexplore.ieee.org*, vol. 64, no. 7, pp. 1105–1119, Nov. 2020, doi: 10.1080/00313831.2019.1649718.

[26] E. B. DONATI, "Educational and Skill Mismatch: An Overview," 2019, Accessed: Feb. 28, 2026. [Online]. Available: https://tesi.univpm.it/bitstream/20.500.12075/4639/1/TESI_Bartolucci Donati Edoardo.pdf

[27] A. Vandeplas and A. Thum-Thysen, *Skills mismatch & productivity in the EU*. 2019. Accessed: Jan. 14, 2026. [Online]. Available: https://www.euroguidance.nl/_images/user/2019/EMPL_Study_Skills mismatch_040719clean.pdf

[28] P. N. Gamage, H. M. A. D. K. Herath, I. Kaluarachchige, and R. A. Hewapathirana, "Aligning Labour Market Demands with a Skilled Workforce: Strategies for Engaging Sri Lankan Youth in TVET Courses," *jhrmp.sljol.info*, vol. 8, no. 1, pp. 1–24, Jun. 2023, doi: 10.4038/jhrmp.v8i1.16.

[29] B. Balasooriya and I. Kulathunga, "A Study to Identify Ways of Enhancing Industry Partnership Towards Technical Education-The Sri Lankan Context," 2020, Accessed: Feb. 05, 2026. [Online]. Available: https://pub.cpsctech.org/steps04/pdf.pdf#page=41

[30] F. Biagi, C. Muñoz, and D. Pietro, "Mismatch between Demand and Supply among higher education graduates in the EU," 2020, doi: 10.2760/003134.

[31] B. A. Gebrehiwot, "Industrial policy and structural transformation in Ethiopia," 2022, Accessed: Feb. 28, 2026. [Online]. Available: https://hal.science/hal-04536958/

[32] D. Rotatori, E. J. Lee, and S. Sleeva, "Fourth industrial revolution skillsets and employability readiness for future job," *academia.edu*, vol. 24, no. 1, pp. 92–103, 2021, doi: 10.1080/13678868.2020.1767453.

[33] M. Helmi *et al.*, "Exploring the Integration of Artificial Intelligence in Technical and Vocational Education and Training (TVET): applications, benefits, challenges,

and future prospects," *beam.pmu.edu.my*, vol. 3, pp. 57–63, 2024, Accessed: Feb. 28, 2026. [Online]. Available: https://beam.pmu.edu.my/index.php/beam/article/view/189

[34] E. Davern, *Trends and new developments in employment services to support transitions in Sub-Saharan Africa, the Middle East and North Africa Regions*. 2020. Accessed: Feb. 28, 2026. [Online]. Available: https://www.econstor.eu/handle/10419/263085

[35] W. Kipngetich and K. M. Ed, "Assessment of integration of competency based education and training approach for quality education: A case of technical and vocational education and training," 2022, Accessed: Feb. 28, 2026. [Online]. Available: http://repository.kisiiuniversity.ac.ke:8080/xmlui/handle/123456789/8534

[36] A. ABTEW Kero and Z. Olana, "The role of big data analytics on improving technical and vocational education outcomes," *Res. Sq.*, 2023, doi: 10.21203/rs.3.rs-3093973.

[37] S. Ra, B. Chin, and A. Liu, "Challenges and opportunities for skills development in Asia: changing supply, demand, and mismatches," 2015, Accessed: Feb. 28, 2026. [Online]. Available: https://www.skillsforemployment.org/sites/default/files/2024-01/wcmstest4_158655.pdf

[38] R. Caroline, "Institutional Preparedness for the Implementation of Competency-Based Education and Training Curriculum in Public Technical Institutions in Mt. Kenya East, Kenya," 2025, Accessed: Feb. 28, 2026. [Online]. Available: http://41.89.31.5/handle/123456789/2145

[39] J. HAKIZAYEZU, "Challenges facing technical and vocational education and training institutions on youth employment in Rwanda: a case of study of Gasabo district," 2022, Accessed: Jan. 28, 2026. [Online]. Available: http://dr.ur.ac.rw/handle/123456789/2054

[40] A. Abtew and A. Endebu, "The role of big data analytics in improving teacher training in developing countries: A literature Review," 2023, doi: 10.21203/rs.3.rs-3111391/v1.

[41] J. Mukundabantu, "Analysis of job vacancies and skills mismatch in Rwanda using

data mining techniques,(2019-2021),” 2023, Accessed: Feb. 28, 2026. [Online]. Available: http://dr.ur.ac.rw/handle/123456789/2665

[42] S. Boopathy and P. S. Kumar, “Predictive analytics with data visualization,” 2022, doi: 10.21203/rs.3.rs-803205/v1.

[43] M. S. Regassa and T. D. Desissa, “Technology integration status in TVET colleges: evidence from Sheger City, Ethiopia,” *Springer*, vol. 4, no. 1, Dec. 2025, doi: 10.1007/s44217-025-00949-x.

[44] World Bank, “Skills development in Ethiopia: Progress and potential.,” 2021, Accessed: Jun. 05, 2025. [Online]. Available: https://openknowledge.worldbank.org/handle/10986/35214

[45] A. Kebede, A. Asgedom, and A. Asfaw, “What works for youth employment in Africa: a review and empirical analysis of the existing youth employment policies and their impact in Rwanda,” *papers.ssrn.com*, vol. 11, no. 1, 2024, doi: 10.1080/2331186X.2024.2406721.

[46] M. Edward, “Internal Efficiency, Leadership Practices and Competitiveness of Technical and Vocational Education and Training Institutions in Nyeri County, Kenya,” 2024, Accessed: Feb. 05, 2026. [Online]. Available: http://41.89.31.5/handle/123456789/2050

[47] M. FLORAH, “IMPLEMENTATION OF COMPETENCY-BASED CURRICULUM POLICY AND LEARNERS’OUTCOMES IN MAKUENI COUNTY, KENYA,” 2024, Accessed: Feb. 05, 2026. [Online]. Available: https://ir-library.ku.ac.ke/bitstreams/5cd4b925-16c5-4601-b3c3-8825fb4e4e44/download

[48] J. Mirieri, “Linking Technical, Vocational Education and Training to Market Demands: A Case of TVET Institutions in Nairobi County,” 2024, Accessed: Jun. 18, 2025. [Online]. Available: https://ir-library.ku.ac.ke/server/api/core/bitstreams/d9bf13eb-3eb3-4bb1-b944-ed289d00793c/content

[49] A. Zary and N. Zary, “Artificial intelligence in technical and vocational education and training: Empirical evidence, implementation challenges, and future directions,” 2025, Accessed: Apr. 22, 2026. [Online]. Available:

https://www.preprints.org/frontend/manuscript/cadce590828df3f67d77ba9bab8fe737/download_pub

[50] E. Gabreyohannes, “Multilevel Logistic Regression Analysis of Correlates of Graduate Employment by Fields of Specialization,” 2024, Accessed: Jun. 18, 2025. [Online]. Available: https://www.naturalspublishing.com/download.asp?ArtcID=28619

[51] L. Mutembei, “The Nexus between Institutional Factors and Development of Employability Skills of Technical Training Institutions Graduates in Meru County, Kenya,” 2024, Accessed: Feb. 05, 2026. [Online]. Available: http://41.89.31.5/handle/123456789/1811

[52] G. M. Motuka, C. Matara, and J. A. Orodho, “Reforms in Technical and Vocational Education and Training (TVET) in Kenya: Focus on Variables That Influence Sustainable Development,” *johnorodho.com*, vol. 15, no. 6, 2024, doi: 10.7176/JEP/15-6-07.

[53] S. Melesse, A. Haley, and G. B. Wärvik, “Bridging the skills gap in TVET: a study on private-public development partnership in Ethiopia,” *Taylor Fr.*, vol. 21, no. 3, pp. 171–186, 2023, doi: 10.1080/14480220.2022.2159854.

[54] E. T. Shiferaw and B. Endeshaw, “Vocational Education and TVET Graduate Employability in Addis Ababa: Mediating Roles of Labour Market Conditions and Personal Attributes,” 2025, doi: 10.21203/rs.3.rs-8262195/v1.

[55] E. Shiferaw, “Vocational Education and TVET Graduate Employability in Addis Ababa: Mediating Roles of Labour Market Conditions and Personal Attributes,” 2025, Accessed: Apr. 22, 2026. [Online]. Available: https://www.researchsquare.com/article/rs-8262195/latest

[56] A. Dimoska, “The role of regional horticulture Technical and Vocational Education and Training in the graduates’ employability and employment outcomes: Insights from Ethiopia,” 2019, Accessed: Jan. 28, 2026. [Online]. Available: https://studenttheses.uu.nl/handle/20.500.12932/32848

[57] A. Kebede, A. Asgedom, and A. Asfaw, “Is linking technical and vocational training with industry a bridge to employability? Lessons from a literature review for Ethiopia,” *Taylor Fr.*, vol. 11, no. 1, 2024, doi:

10.1080/2331186X.2024.2406721.

[58] S. Melesse, A. Haley, and G. B. Wärvik, “Bridging the skills gap in TVET: a study on private-public development partnership in Ethiopia,” *Taylor Fr.*, vol. 21, no. 3, pp. 171–186, 2023, doi: 10.1080/14480220.2022.2159854.

Appendix

Appendix A: Sample Dataset

	Year	Sector	Skill	Vacancies	Qualification_Requirement	Experience_Years	Graduates	Field_of_Study	Employment_Status	Wage_Level_ETB	Demand_Level
0	2022	ICT	Software Developer	49	Degree	3	40	ICT	Employed	6860	Low
1	2023	Manufacturing	Technician	35	Degree	1	171	Manufacturing	Unemployed	6466	Low
2	2024	Agriculture	Farm Extension Officer	75	Certificate	4	479	Construction	Self_Employed	17363	High
3	2020	ICT	Data Analyst	57	Diploma	1	207	ICT	Employed	16583	High
4	2022	Construction	Electrician	80	Degree	3	326	Manufacturing	Unemployed	12420	High
...	...	...	...	...	...	...	...	...	...	...	...
2445	2025	Services	Customer Service Agent	45	Degree	3	76	Electrical	Self_Employed	11261	Low
2446	2025	Manufacturing	Welder	89	Certificate	4	161	ICT	Self_Employed	15211	High
2447	2023	Manufacturing	Quality Control Officer	64	Certificate	2	306	Electrical	Self_Employed	16268	Moderate
2448	2021	Services	Hotel Receptionist	23	Degree	1	86	Agriculture	Employed	17957	Low
2449	2022	Manufacturing	Quality Control Officer	80	Diploma	0	198	ICT	Unemployed	6496	High

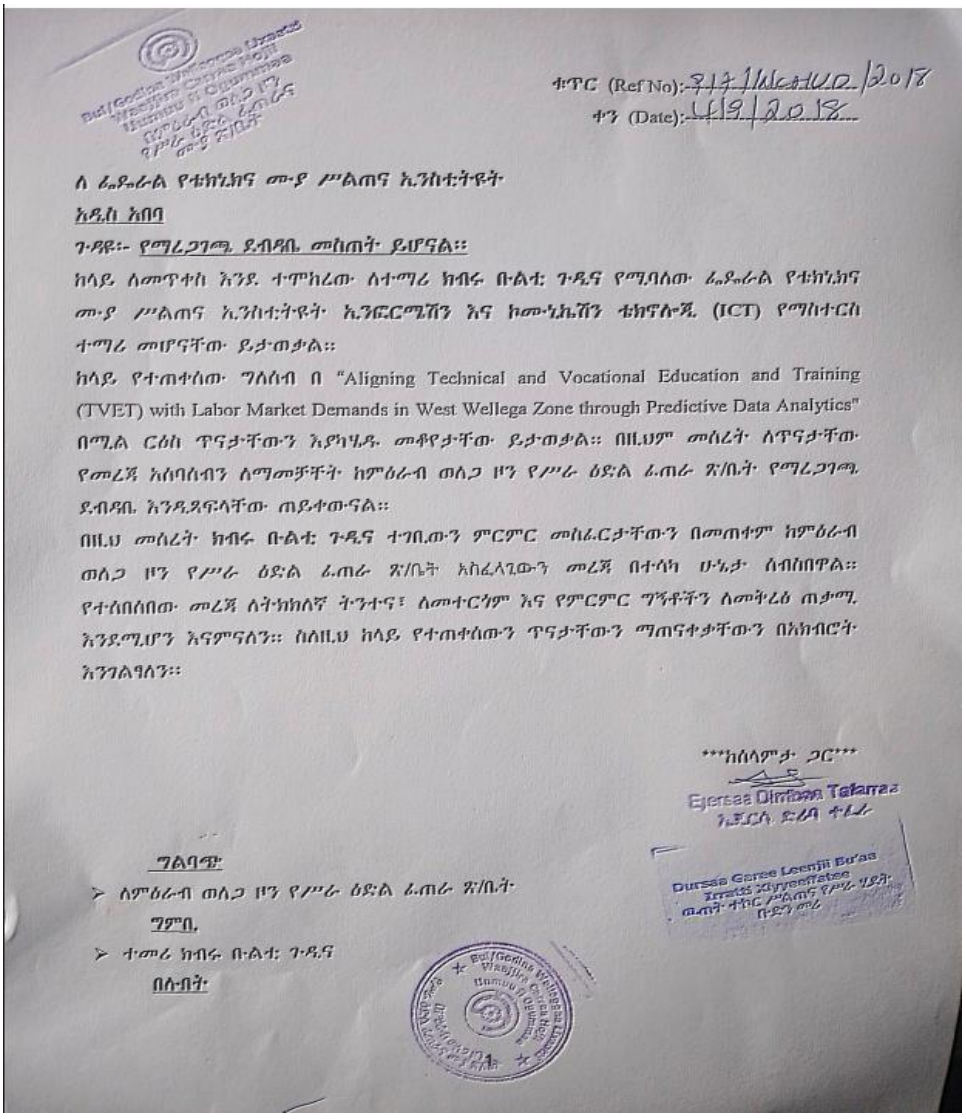
2450 rows × 11 columns

Appendix B: Data Dictionary

Table 19: Complete Data Dictionary for Labor Market Dataset

Variable Name	Description	Data Type	Allowed Values/Range
record_id	Unique identifier for each observation	Integer	1 - 2,450
year	Calendar year of data collection	Integer (Discrete)	2020, 2021, 2022, 2023, 2024, 2025
sector	Economic sector/industry category	Categorical (Nominal)	Construction, Agriculture, ICT, Manufacturing, Services
occupation	Specific job title or skill category	Categorical (Nominal)	45 unique values
vacancies	Number of job openings in sector/occupation	Integer (Continuous)	10 - 100
graduates	Number of TVET graduates in the corresponding field	Integer (Continuous)	20 - 499
wage_level	Average monthly salary in Ethiopian Birr (ETB)	Float (Continuous)	6,006 - 17,992 ETB

Appendix C: Ethical Approval Letter



ቁጥር (Ref No): ጸ፲፯/ዘሬ፰ዘሀ.ዐ/2018
ቀን (Date): 4/9/2018

ለ ፌዴራል የቴክኒክና ሙያ ሥልጠና ኢንስቲትዩት
አዲስ አበባ

ጉዳዩ፡- የማረጋገጫ ደብዳቤ መስጠት ይሆናል፡፡

ከላይ ለመጥቀስ እንደ ተሞከረው ለተማሪ ክብሩ ቡልቲ ጉዲና የሚባለው ፌዴራል የቴክኒክና ሙያ ሥልጠና ኢንስቲትዩት ኢንፎርሜሽን እና ኮሙኒኬሽን ቴክኖሎጂ (ICT) የማስተርስ ተማሪ መሆናቸው ይታወቃል፡፡

ከላይ የተጠቀሰው ግለሰብ በ "Aligning Technical and Vocational Education and Training (TVET) with Labor Market Demands in West Wellega Zone through Predictive Data Analytics" በሚል ርዕስ ጥናታቸውን እያካሄዱ መቆየታቸው ይታወቃል፡፡ በዚህም መሰረት ለጥናታቸው የመረጃ አሰባሰብን ለማመቻቸት ከምዕራብ ወለጋ ዞን የሥራ ዕድል ፈጠራ ጽ/ቤት የማረጋገጫ ደብዳቤ እንዲጻፍላቸው ጠይቀውናል፡፡

በዚህ መሰረት ክብሩ ቡልቲ ጉዲና ተገቢውን ምርምር መሰረታቸውን በመጠቀም ከምዕራብ ወለጋ ዞን የሥራ ዕድል ፈጠራ ጽ/ቤት አስፈላጊውን መረጃ በተሳካ ሁኔታ ሰብስበዋል፡፡ የተሰበሰበው መረጃ ለትክክለኛ ትንተና፣ ለመተርጎም እና የምርምር ግኝቶችን ለመቅረፅ ጠቃሚ እንደሚሆን እናምናለን፡፡ ስለዚህ ከላይ የተጠቀሰውን ጥናታቸውን ማጠናቀቃቸውን በአክብሮት እንገልፃለን፡፡

ከሰላምታ ጋር
Ejersaa Dimtsaa Tafarraa
አደርሳ ድሪባ ተፈራ

Durssa Gerse Leenjii Bu'aa Imaltti Xiyyeeffatee
ዉጤት ተኮር ሥልጠና የሥራ ዕደላ ቡድን መሪ

ግልባጭ
- ለምዕራብ ወለጋ ዞን የሥራ ዕድል ፈጠራ ጽ/ቤት
  ግምቢ
- ተመሪ ክብሩ ቡልቲ ጉዲና
  በሱቢት