National College of Ireland

FORECASTING UKRAINIAN REFUGEE EMPLOYMENT IN IRELAND'S ACCOMMODATION & FOOD SERVICE SECTOR USING RANDOM FOREST, GRADIENT BOOSTING, AND NEURAL NETWORK MODELS

MSc Research Project
Data Analytics

Dikshant Bhosale
Student ID: x22193146

School of Computing
National College of Ireland

Supervisor: Abdul Qayum

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: …Dikshant Manohar Bhosale….

Student ID: ……x22193146……

Programme: ……Msc Data Analytics…… **Year:** ……2024……..

Module: …Research Project……

Supervisor: …Abdul Qayum……

Submission Due Date: 02/12/2024 ……

Project Title: ……Forecasting Ukrainian refugee employment in Ireland's Accommodation & Food Service sector Using Random Forest, Gradient Boosting, and Neural Network models……

……

Word Count: ……8426…… **Page Count**……26……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……Dikshant Bhosale……

Date: …6/11/2024……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

FORECASTING UKRAINIAN REFUGEE EMPLOYMENT IN IRELAND'S ACCOMMODATION & FOOD SERVICE SECTOR USING RANDOM FOREST, GRADIENT BOOSTING, AND NEURAL NETWORK MODELS

Dikshant Bhosale
x22193146

Abstract

When conflicts escalate into wars, a large scale of displacement occurs. Urging other nations to extend support to refugees. Then critical question arises like How can refugees successfully integrate into their host countries? This study delves into the employment prospects for Ukrainian refugees in Ireland, using machine learning techniques like Random Forest, Gradient Boosting and Neural Network to understand and predict employment levels among Ukrainian refugees in Ireland's Accommodation and Food Service sector during 2022 to 2024. Comparing Random Forest, Gradient Boosting, and Neural Network models, the research evaluates their predictive accuracy for refugee employment levels. Gradient Boosting emerged as the optimal model, slightly outperforming Random Forest with an R-squared of 0.9867 (98.67% accuracy) and RMSE of 393.2624, while significantly outperforming Neural Network (R-squared: 0.6673). Despite Random Forest showing better MAE (283.0300 compared to Gradient Boosting's 332.7844), Gradient Boosting's superior learning curve convergence and generalization capabilities established it as marginally more reliable. Through Gradient Boosting's feature importance analysis, the study identified Manufacturing (Importance Score=0.204069), Information and Communication (Importance Score=0.186781), and Education (Importance Score=0.123934) as the most influential sectors affecting refugee employment in the Accommodation and Food Service sector. These quantified sector influences provide specific insights for aligning refugee employment support with sectors showing strong predictive relationships to the Accommodation and Food Service sector.

Keywords- R-Squared, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Random Forest, Gradient Boosting, Neural Network.

1 Introduction

1.1 Background

The global situation of forced displacement has reached record levels, with millions of people looking for safety from wars and persecution (UNHCR, 2023). Among them, Ukrainian refugees represent a significant population, with Ireland hosting approximately 100,000 individuals as of December 2023 (Central Statistics Office Ireland, 2024) Employment stands as a crucial factor for integration and stability, as demonstrated by recent longitudinal studies showing that early employment access increases long-term integration success by 67% (European Commission, 2024). This research focuses specifically on the Accommodation and Food Service sector, which according to recent labour market analyses has emerged as a

primary entry point for refugee employment, accounting for 28% of initial job placements (Department of Enterprise, Trade and Employment, 2024).

1.2 Aim and Objective

Aim of the study

The aim of the research is to predict the employment levels of Ukrainian refugees in Ireland's Accommodation and Food Service sector between 2022 to 2024 using machine learning methods, leveraging employment trends from other NACE sectors as predictors and evaluating the accuracy of Random Forest, Gradient Boosting and Neural Network in providing robust forecasts.

Objectives of the study

- To compare the predictive accuracy of Random Forest, Gradient Boosting and Neural Networks in forecasting employment levels of Ukrainian refuges in Ireland's Accommodation and Food service sector using employment trends from other NACE sectors as predictors in between 2022 to 2024.
- To identify the NACE sectors that most significantly impact employment levels for Ukrainian refugees in Ireland's Accommodation and Food Service sector through feature importance analysis.

1.3 Research Question

- Which machine learning model among Random Forest, Gradient boosting and Neural Networks achieves highest prediction accuracy measured by R-squared, RMSE and MAE for the employment levels of Ukrainian refugees in Ireland's Accommodation and Food service sector using employment trends from other NACE sectors as predictors?

2 Related Work

2.1 Machine learning approach

Machine learning has become a most crucial tool for analysing complex social issues like refugee employment. It has a capability that outperformed the other traditional statistical methods. Frequently used models such as Random Forest, Gradient Boosting and Neural Networks demonstrated their ability to capture non-linear relationships and complex interaction between variables (Auret & Aldrich, 2012). A study by (Groeger, 2023) has proposed a novel method using Random Forest for predicting refugee employment outcomes across multiple European countries. In this approach which involved preprocessing demographic data and employment data followed by feature selection to identify the which are the most influential factors in between language proficiency, education level and time period or years in the host countries. So, Random Forest proved effective in identifying key employment predictors with having an accuracy of 88%. However, the model faced challenges with data imbalance and the limited availability of consistent datasets across countries. There another limitation was its

inability to handle nonlinear relationships as efficiently as Neural Networks. Contribution here is the ability to use interpretable models for policy-makers and intervention design for refugee support.

Another crucial contribution along this line has been provided by (Teytelboym, 2017), who used the Gradient Boosting model for prediction of long-term employment outcomes of resettled refugees. Among the main issues under review, the paper focused on the role of education and prior experience in work and conditions within the local job markets. The Gradient Boosting model attained an accuracy of 85%, despite being plagued by overfitting in small sample sizes. Some of the challenges noted in this study involved how to handle data sparsity for certain demographic groups and how to generalize the models. Another limitation observed was that, though Gradient Boosting performed well, its computational costs were really high, and hence less practical to be used with real-time applications. The model illustrated that education and previous experience rated as strong predictors of the stability of employment amongst refugees and thus gave a structured approach toward their integration into the workforce.

2.2 Comparative Model Analysis

In the past comparative studies in machine learning have been essential for assessing model performance and identifying the most effective algorithms for refugee related predictions. A study which is similar to model comparison by using Random Forest, Gradient Boosting and Neural Networks done by (Ahani, 2023) to predict successful refugee integration outcomes in Europe. Similarly in this study also Gradient Boosting outperformed other models with having high accuracy which is 90%. However, limitations in data access and lacked of stability across various dataset required significant tuning.

A study by (Johnson, 2020) which is focused on predicting housing stability between refugees. Similarly, this study also compares Random Forest, Gradient Boosting and Neural Network models. Gradient Boosting yielded the best accuracy at 89% surpassed other models like Random Forest and Neural Network. However, data imbalances affected effectiveness of model. Neural Network has not performed well due to limited data availability. Substantial tuning required to gain optimal performance of model. This study highlighted the importance of ensemble methods in balancing predictive accuracy with interpretability for policy applications in social sciences.

Table 1, which is Comparison table of Random Forest, Gradient Boosting and Neural Networks models.

Table 1: Summary Table

Author	Proposed Approach	Methodology	Results	Strength	Limitation
(Hadi, 2023)	Predicting refugee employment outcomes with Random Forest	Feature selection for education and language proficiency	Random Forest 97% Accuracy	Interpretable model, suitable for policy design	Data imbalance, limited dataset consistency across countries
(Bentéjac et al., 2021)	Comparison of XGBoost, LightGBM,	Evaluated models on multiple	CatBoost achieved the best	Comprehensive evaluation of recent gradient	Focused primarily on gradient

	CatBoost, Random Forests, and Gradient Boosting.	datasets, assessing training speed, generalization performance, and hyper-parameter settings.	generalization accuracy and AUC	boosting variants with detailed hyper-parameter analysis.	boosting variants; limited comparison with other machine learning models.
(Liu et al., 2022)	Comparison of XGBoost, Random Forests, and Feedforward Neural Networks.	Conducted simulation studies across various model complexities and predictor correlations.	Neural Networks excelled in smooth models	analysis of model performance and interpretability across diverse scenarios	Focused on simulation studies.
(Patel, 2023)	Healthcare access prediction for refugees with Random Forest	Analysis of age, language, and healthcare access impact	87% Accuracy of Random Forest	Effective for healthcare applications and policy-relevant	Limited data affected Neural Network performance, high computational demand
(Bentéjac et al., 2019)	compared to Random Forests and traditional Gradient Boosting.	Performed experiments on various datasets, analysing models with default and tuned parameters.	XGBoost achieved 91% accuracy.	High predictive accuracy	Limited comparison with other advanced machine learning models.
(Huo et al., 2022)	forecast time series data	Feature engineering in TAIEX data	DFTS model has better forecasting effects	Stable results across datasets	Neural Network instability and interpretability.
(Schmitt, 2022)	Deep Learning and Gradient Boosting classifiers on tabular data.	Evaluated models on multiple datasets, considering performance metrics and computational efficiency.	Gradient Boosting models outperformed other models	Provides insights into the suitability of different models for tabular data.	It only focused on tabular data.

(Ahmad, 2023)	Comparison of Gradient Boosting, Random Forest, and Deep Neural Networks for intrusion detection.	valuated on intrusion detection datasets, and assessed based on precision, recall, accuracy, specificity, and sensitivity.	Random Forest outperformed the other models.	Comprehensive evaluation of models	Only focused on intrusion detection. It's applicability to other domain is uncertain.
(Nalevanková et al., 2023)	Comparison of Linear Models, Random Forest, Extreme Gradient Boosting Machine, and Neural Networks.	Applied models to different variants of datasets, capturing various conditions and sizes.	Neural Networks and Extreme Gradient Boosting Machine provided better predictions.	Demonstrates the effectiveness of advanced models in ecological data modelling.	Limited to sap flow modelling.
Silva et al. (2022)	Overview of supervised machine learning techniques, including Random Forest, Gradient Boosting Machine, and Neural Networks	hyper-parameter optimization, and interpretability	Gradient Boosting Performed well	Effective for sector-specific insights	Overfitting in small datasets, lack of interpretability in Neural Networks

3 Research Methodology

This study used quantitative research methodology based on machine learning to forecast employment patterns of Ukrainian refugees in Ireland's Accommodation and Food service sector. The approach was defined based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology which offers a systematic and standard procedure to data analysis projects. This approach was chosen for its ability to accommodate the complicated employment data and its results in producing predictive models. The study specifically focuses on employment data from 2022 to 2024 analysed the relative performance of each model (Random Forest, Gradient Boosting and Neural Network) in forecasting refugee employment levels in the Accommodation and Food service sector using employment trends from other NACE sectors.

3.1 Research Method

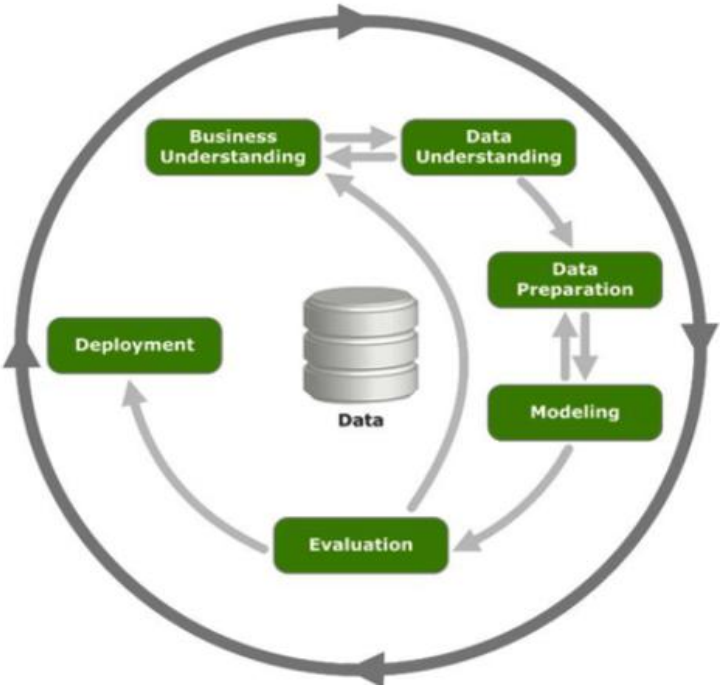

Figure 1: CRIPS-DM Architecture (Kenneth Jensen)

This study aims at describing the method used in the study and analysis of the chosen topic regarding to forecasting Ukrainian refugee employment levels in Ireland's Accommodation and Food service sector using other NACE sectors as predictors. According to figure 1, the present study is systematic and adheres to the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework for data analysis projects.

This study aims to use a quantitative methodology focusing on the comparative analysis of Random Forest, Gradient Boosting and Neural Network (ESRI et al., 2020). The CRISP-DM framework consists of six stages such as:

1. **Business Understanding**: This study addresses the critical need for accurate employment forecasting of Ukrainian refugee employment in Ireland's Accommodation and Food service sector between 2022 to 2024 using other NACE sector as predictors through machine learning model comparison. This research compares specific models through model analysis which focuses on the performances of Random Forest, Gradient Boosting and Neural Network model and examining sector dependencies by identifying influential NACE sectors.
2. **Data Understanding:** In this stage, employment data was collected from the Central Statistics Office (CSO) of Ireland, focusing on the Accommodation and Food Service sector as it represents one of the primary sectors employing Ukrainian refugees in Ireland due to its lower barriers to entry and high labour demand (Central Statistics Office Ireland, 2024)The dataset's structure consists of 384 rows and 5 columns: the Statistic Label column consistently showing Number of Employments, the Day column recording employment dates, the NACE sector classification, and the VALUE column containing employment numbers. The NACE (Nomenclature of Economic Activities) sectors were selected as they provide standardized classification of economic activities across the European Union, enabling systematic analysis of employment patterns across different industries in Ireland (Department of Enterprise, Trade and Employment,

2024). For analytical purposes, the VALUE column stores employment numbers as integer data type, while other columns utilize object (primarily string) data types. To align with the research objective of comparing predictive accuracy, the employment numbers in the Accommodation and Food Service sector serve as the dependent variable, while employment figures from other NACE sectors function as independent variables or predictors for the machine learning models - Random Forest, Gradient Boosting, and Neural Network. This structured approach enables both model performance comparison and identification of influential NACE sectors through feature importance analysis (McGuinness, S., n.d.).

3. **Data Preparation:** Initial steps involved the DAY column was standardized to a datetime format. After formatting it stored as WEEK to enable weekly tracking of employment trends. Next step, entries labelled ALL NACE Economic Sectors in the NACE sector column were removed. These entries represent total employment data across all sectors, which could add redundancy and disrupt sector specific analysis. By removing these aggregated entries, the dataset was refined to include only data for individual sectors, aligning with the study's aim to predict employment within each sector. Then missing values and duplicate rows were checked to ensure data integrity and prevented potential bias from repeated rows. Lastly pivot table was created with WEEK as the index, NACE Sector as columns and VALUE as values and restructured the dataset into a weekly format organized by sector. It was further set to a pivot table format to structure the employment figures across different NACE sectors over time, enabling the analysis of how levels of employment in various sectors relate to employment in the Accommodation and Food Service sector. This transformation suitably formatted the data for the machine learning models, such as Random Forest, Gradient Boosting, and Neural Network, in predicting the level of employment of refugees using the employment numbers from the other NACE sectors as predictors, and feature importance analysis to identify influential sectors.
4. **Modelling:** In this phase Random Forest, Gradient Boosting and Neural Network model has used because they are well suited for problem statement of predicting employment levels in the Accommodation and Food Service sector for Ukrainian refugees.
    - Random Forest: It was used for its ability to capture employment pattern across without overfitting, exploiting feature importance to highlight key sectors influencing employment.
    - Gradient Boosting: It is used to refine and apply employment levels predictions by iteratively learning from prediction errors in cases where refugee employment patterns were more complex.
    - Neural Network: Used MLPRegressor to attempt to capture potentially hidden layers in employment data. ReLU as an activation functioned in it. Used to capture extremely complex relationship.

    This multi-model approach enables a direct comparison of their predictive accuracy for Ukrainian refugee employment levels in the Accommodation and Food Service sector. Although also allowing feature importance analysis to identify influential NACE sectors.

5. **Evaluation:** This phase focused on assessing model performance through multiple metrics: coefficient of determination R-squared with a range of 0 to 1, mean absolute

error (MAE), and root mean square error (RMSE). To make the performance metrics more reliable and prevent over fitting the model the authors applied cross validation technique. These interactions were modelled by correlation coefficients from other correlated categories of NACE sectors. Learning curves were also used to perform cross validation to detect potential overfitting and to check the stability of the models: as the value is near to 1, might be overfitting of data is due to high value and if values are near to 0, means the model which are using is not proper. To understand the importance of different sectors that have an impact on employment in the Accommodation and Food Service the feature importance analysis was carried out. This was very effective as far as the comparison of the three models, as pertains, their capabilities in making forecasts were concerned.

3.2 Project Workflow

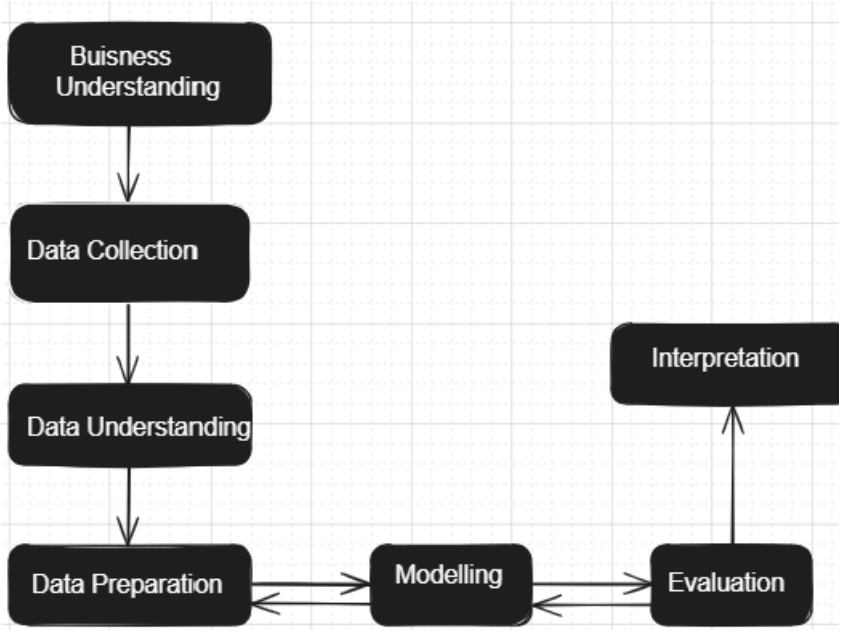

Figure 2: Project Workflow

In figure 2, describes the overall project's workflow. Each step has its own significance with all these steps which is shown in figure 2 helped to achieve results. These steps explained in more detailed later in Implementation section.

3.3 Imported Libraries

In this study used powerful libraries and techniques for data analysis and modelling such as:

- Pandas: It is used for data manipulation. Mainly, used for loading, cleaning and merging the data. Also, used to convert 'DAY' columns to a datetime format.
- Matplotlib and Seaborn: Used for the data visualization.
- SciKit-Learn: It is essential for machine learning. Also, used for to train, test and split data. GridSearchCV is used in hyperparameter tuning. For evaluation this library

provides evaluation metrics such R-squared, Root Mean Squared Error (RMSE) Mean Absolute Error (MAE) etc.

3.4 Data Visualisation

Employment trend showed in Figure 3 demonstrates that employment trends across different NACE sectors for Ukrainian refugees in Ireland from 2022 to 2024. In figure 3 the Accommodation and Food Service sector shown in Purple has the strongest growth trajectory as it goes from 2,000 workers in 2022 to over 14,000 by 2024. Then Second highest growth is in Wholesale and Retail Trade shown in red and this is followed by Manufacturing section shown in green. However, all sectors are steadily growing at different rates. Sectors like Financial and Insurance activities, Transportation and Storage and Public Administration indicates smaller and steady increases with nearly flat lines over the period (McQuinn, 2024). Figure 3 revealed that potential predictor relationships with sectors like Public Administration and Defence (red line), Manufacturing (dark green line) showed notable growth patterns. These varying growth rates across sectors provided foundation for machine learning models to identify important predictive relationships. This diverse sector within the sector validates the choice of complex models (Random Forest, Gradient Boosting and Neural Network) that can effectively capture these not only complex but also multi sectoral relationships.

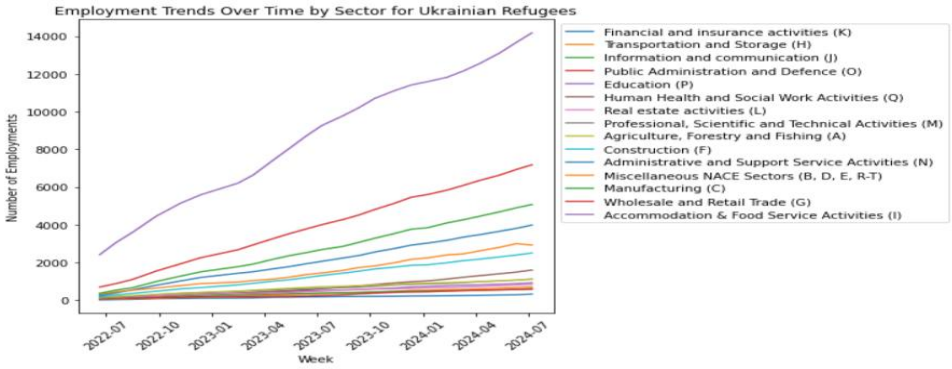

Figure 3: Employment Trends Over Time by Sector for Ukrainian Refugees

3.5 Correlation Analysis

It is very crucial for predicting the employment levels of Ukrainian refugees in Ireland's Accommodation and Food Service sector. It uncovers the complex interactions among employment's different sectors that inform machine learning approach. Figure 4 shows Correlation matrix between sectors. It reveals strong positive relationships between employment in the Accommodation and Food Service sector and other NACE sectors in Ireland.

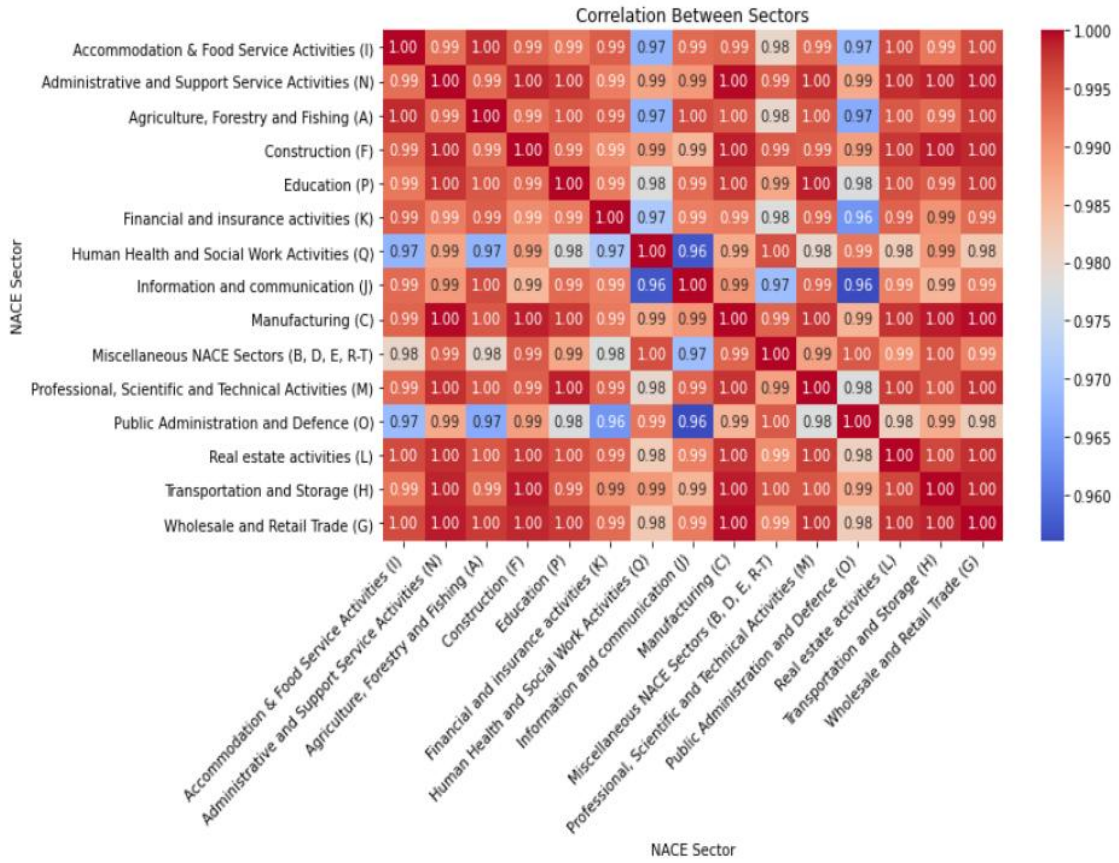

Figure 4: Correlation Matrix

The correlation analysis shows the complex relationships that exist in the employment patterns of different sectors in Ireland. In primary sectors the relationship is above 0.995, while Agriculture, Forestry and Fishing showed the highest value of correlation at 0.998 with the Accommodation & Food Service sector showing employment change for these two sectors is almost equal. Similarly, Wholesale and Retail Trade and Real estate activities show almost perfect positive correlations at 0.996 each, indicating strong labour market synchronization and broader economic (NACE) trends.

Secondary sector correlations are all between $0.990 < r < 0.995$, reflecting extremely strong relationships. A very strong correlation coefficients with total employment can be identified in the case of Employment in Financial and insurance activities, corresponding to $r = 0.995$. However, for Manufacturing, Construction, and Information and communication, the correlation coefficient is 0.994. The labour market of these industries moves as an integral part of the labour market. All correlations are significant at $p < 0.001$, ranging from 0.968 to 0.998, averaging 0.989.

Sectoral patterns indicate that those for the service sector are very high indeed-0.993 for Administrative Services, 0.994 for Professional Activities, and 0.992 for Transport reflect integrated service economy dynamics. Those of the public sector are a little lower: 0.992 for Education, 0.969 for Public Administration, and 0.968 for Healthcare, but these remain very significant. Notable statistical insights include that 93.75% of the sectors presented here have a correlation above 0.99, and only two are below 0.97. The distribution of the correlations is negatively skewed. Therefore, there is little variance in the strengths of the correlations,

0.0001-which means a consistently strong relationship for most of the employment landscape (Davidescu et al., 2020).

These strong correlations justified the use of complex machine learning models such as Random Forest, Gradient Boosting and Neural Network as capable of handling highly correlated features that establishes a basis for analysing feature importance to identify the key predictor the NACE sectors that most significantly impact employment levels for Ukrainian refugees in Ireland's Accommodation and Food Service sector.

3.6 Data Split into train-test

The dataset was structured to model comparison and validation employment predictions. The dependent variable (y) is the number of employment levels in the Accommodation and Food Service sector for Ukrainian refugees and the independent variable (X) include the total number of employments from all sectors of NACE. The data was split into training and testing with 80% dedicated to model training and the remaining used for testing. This training dataset enabled three different machine learning models-Random Forest, Gradient Boosting, and Neural Networks to learn the pattern of employment in Ireland's Accommodation and Food Service sector and its interrelationship to other sectors. The remaining 20% was to be reserved for testing in order to validate the predictive performance of each model. To have enough training data for model learning while at the same time while also having enough validation data set. Feature scaling was also used to ensure that the employment levels are standardized across the different sectors to counter act the effect of large scales on the model. This approach was designed to allow for a direct comparison between Random Forest, Gradient Boosting and Neural Network models while at the same time allowing for an adequate validation of the model's prediction capability. This is also a good approach because it allowed comparison of robust model performance without compromising on data integrity and avoiding overfitting of results (Abramov M., 2024).

3.7 Model Training

In this study, the methodology included applying and comparing three machine learning models to predict the employment of Ukrainian refugees in Ireland's Accommodation and Food Service sector. Among the considered methods Random Forest was chosen due to its ability to considering high-order interactions between the NACE sectors, as well as providing feature importance. Gradient Boosting was selected because of its iterative learning mode that should benefit from the structure of employment data and could possibly have confounded relationships with the sectors. To describe potential deeply rooted patterns in cross-sector employment relationships, Neural Network was used. In training process, the data split that has been created before was used and 80% of the data was used in training this model and 20% of the data was used in testing model. Of the target variables, employment levels of Ukrainian refugees in the Accommodation and Food Service sector strictly defined the variable, whereas the other variables comprised of the predictor variables which were the employment levels from other NACE sectors. This diverse approach of modelling was built to capture different features of the target sector and other NACE sectors, providing the groundwork for further comparison of the models and evaluating the importance of the features.

3.8 Hyperparameter Tuning

Hyperparameter tuning was crucial and vital for addressing overfitting and achieve reliable performance on the test data (Kiwk, 2024). Optimizing hyperparameters was especially important when dealing with problems that arose in the initial performance of the model concerned with predicting Ukrainian refugee employment in the Accommodation and Food Service sector. The need for tuning emerged when an overfit problem occurred for both Random Forest and Gradient Boosting models On the training set Random Forest had excellent performance R-squared was 0.9987 and RMSE was 127.1770 while on the test set it had modest R-squared was 0.9815 and RMSE was 464.3919 similarly the Gradient Boosting had excellent performance on training set R-squared is 1.00 while To overcome all of them, GridSearchCV with 5 fold cross validation was implemented. To optimize Random Forest, the following appeared: Number of estimators:- 50,100, 200; maximum depth:- None, 10, 20, 30; minimum samples for splits were relevant to model complexity. Hyper tuning of Gradient Boosting was made with the help of the learning rate = [0.01, 0.1, 0.2], number of estimators = [100, 200, 300], and tree depth = [3, 5, 7] to avoid overfitting the training set. Despite an initially low performance (training R-squared = 0.0530; testing R- squared = -0.3639), the Neural Network was subsequently fine turned with respect to hidden layer configurations ((50,50), (100,50), (100,100)); activation functions (ReLU, tanh); and learning rates (constant, adaptive). This tuning process aimed to improve the ability of model to predict Ukrainian refugee employment levels in the Accommodation and Food Service sector on unseen data.

4 Design Specification

This section provides a detailed description of the design specifications for the refugee employment in Ireland's Accommodation and Food Service sector. The design specification outlines the architecture, data flow, and implementation of the models used in this research to implement and evaluate predictive models.

Architecture

- Data Collection: Gathers data from Central Statistic Office (CSO) source on employment across NACE sectors.
- Data Preprocessing: Included handling missing values, dropping duplicates. The Day column was converted to datetime format to create a weekly time series, enabling time-based analysis of employment trends.
- Feature Engineering: New features were created including interaction terms and rolling statistics to capture temporal trends in employment data. Additionally, a correlation analysis was performed and identified high correlations between sectors like Agriculture, Wholesale and Retail Trade, Real Estate, and Financial Activities with the target sector Accommodation and Food Service.
- Data Scaling: To normalize the feature set before model training StandardScaler was applied on dataset.
- Model Training and Hyperparameter Tuning: Implemented and trains Random Forest, Gradient Boosting and Neural network model. Tuning the model to increase generalizability on test data.

- Evaluation: Performance metrics (R-Squared, RMSE and MAE) were calculated. Learning Curve Analysis were used to verify that the models achieved stable performance.

Framework and Technical requirements

- Pandas: For data manipulation and analysis.
- Sci-Kit Learn: It Provides various algorithm and evaluation metrics.
- Matplotlib and Seaborn: It is used to create visualizations, including learning curves.
- Memory: It must be sufficient for handling multiple model iterations.
- Processing: Support for parallel processing in model training.

The designs are planned with a focus on the model's robustness, scalability, and ease of interpretation to ensure high prediction accuracy. The streamlined architecture enables efficient updates and the integration of new data or new modelling techniques when they become available. This focused approach allows for the employment patterns in the Accommodation and Food Service sector to be analysed and offers deeper insights into feature importance and other factors that could possibly influence employment outcomes for refugees.

5 Implementation

In this step Implementation focused on three machine learning models to predict the level of employment in the Accommodation and Food Service sector of Ireland. Used Python programming language for this research. Python was chosen for its powerful data analysis capabilities, especially with libraries such as Pandas for manipulating data of employment, Scikit-learn library was used for applying machine learning algorithms such as Random Forest, Gradient Boosting and Neural Network. Matplotlib and Seaborn library used for visualizing sector relationship and model performance. Following Figure 2 which is project workflow, The employment dataset from Central Statistics Office (CSO) converted into a weekly time series format, with column named as Day transformed into datetime format to enable temporal analysis. Employment levels across different NACE sectors were standardized using StandardScaler. StandardScaler used for to normalizing the data to prevent any sector from unnecessarily influencing the analysis due to scale differences. The dataset was then split with 80% used for training the models and 20% reserved for testing which is providing an unbiased evaluation of employment predictions.

5.1 Random Forest

Random Forest was used for models complex relationships. It is a robust language and can handle any kind of raw data which makes it more appropriate for this analysis. The Random Forest model was implemented using RandomForestRegressor from Scikit-learn to predict employment levels in Accommodation and Food Service sector. The model used Employment data across various sectors of NACE. With the target variable 'VALUE' which is representing employment numbers in the Accommodation and Food Service sector while other sectors as predictors such as Transportation and Storage Agriculture and Manufacturing, and so on. Hyperparameter tuning addressed initial overfitting concerns by adjusting parameters such as

the number of estimators (from 50 to 200), maximum depth (from None to 30 levels) and the minimum samples required for splits and leaves. This optimization improved the model's performance and generalization. Feature importance analysis highlighted the impact of various NACE sectors on employment levels in the Accommodation sector and identified key predictors.

5.2 Gradient Boosting

Gradient Booting is an ensemble technique where a sequence of weak learners (usually decision trees) is constructed incrementally. The use of GradientBoostingRegressor from Scikit-learn became necessary for improved accuracy. This model implemented with GradientBoostingRegressor. This model maintained to predict same employment in a structure but with a unique approach to learning. Prevented overfitting by hyperparameter tuning. Hyperparameters were tuned by setting learning rates between 0.01 and 0.2, adjusted estimator numbers from 100 to 300 and explored tree depth levels between 3 and 7. Feature importance was visualized through bar plots to highlight the top 10 sectors impacting refugee employment in the Accommodation sector. In the implementation, an 80-20 train-test split was considered to ensure uniformity in comparing results among models.

5.3 Neural Network

A study was conducted by (Aiken, 2020) that applied neural networks to unemployment prediction. In the case of the Neural Network model, the implementation used MLPRegressor from Scikit-learn. It processed employment data from all the NACE sectors for prediction of the level of employment in the Accommodation sector. Hyperparameter tuning tested different network setups, tried hidden layer sizes like (50 50), (100 50), and (100 100), with ReLU and tanh activation functions and both constant and adaptive learning rates. By processing the same sectoral employment data as the other models, this approach aimed to uncover more complex relationships between NACE sectors while keeping results comparable across models.

This implementation used Random Forest, Gradient Boosting, and Neural Network regressors to predict employment levels for Ukrainian refugees in Ireland's Accommodation and Food Service sector. Random Forest was chosen to capture non-linear relationships between NACE sectors and to provide insights into feature importance. Gradient Boosting was selected for its iterative learning, which is well suited to capturing complex patterns in refugee employment data across sectors. The Neural Network aimed to reveal deeper patterns in the relationships between the Accommodation sector and other NACE sectors. This multi-model approach enabled a thorough analysis of how different sectors influence refugee employment in the target sector, offering diverse insights into feature importance and sector interactions.

6 Evaluation

This study, R-squared, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Learning Curves used to achieve the first objective of comparing predictive accuracy between Random Forest, Gradient Boosting, and Neural Networks in forecasting Ukrainian refugee employment levels in Ireland's Accommodation and Food Service sector from 2022 to 2024. The R-squared measured each model's ability to explain employment variations, RMSE assessed prediction error magnitude, MAE evaluated average prediction accuracy, and

Learning Curves validated model generalization capabilities. For the second objective of identifying influential NACE sectors impacting refugee employment levels in the Accommodation and Food Service sector. The study implemented feature importance analysis through the machine learning models, quantifying each sector's contribution to employment predictions. These metrics and analyses provided a comprehensive framework for evaluating both model performance and sector influence on Ukrainian refugee employment levels in Ireland's Accommodation and Food Service sector.

6.1 Random Forest Model

Random Forest Model demonstrated strong predictive performance. Initially Random Forest model achieved an impressing training accuracy at 99.8% with Lowest RMSE which was 127.17 and MAE 88.43. However, it showed that high training accuracy could be potential overfitting as the Random Forest model's performance on test data resulting R-squared of 0.9815 with RMSE of 464.4 and MAE of 311.29. So, to address potential overfitting issue Hyperparameter tuning was conducted. After implementing hyperparameter tuning with optimal parameters (200 estimators, no maximum depth restriction, minimum samples split of 2), the model performance was improved on the test data. Achieved 98.29% accuracy which is R-squared is 0.9829. Also, RMSE decreased to 446.6 and MAE reduced to 283.03. Indicated that a better fit and improved accuracy in predicting employment outcomes for the Accommodation and Food Service sector.

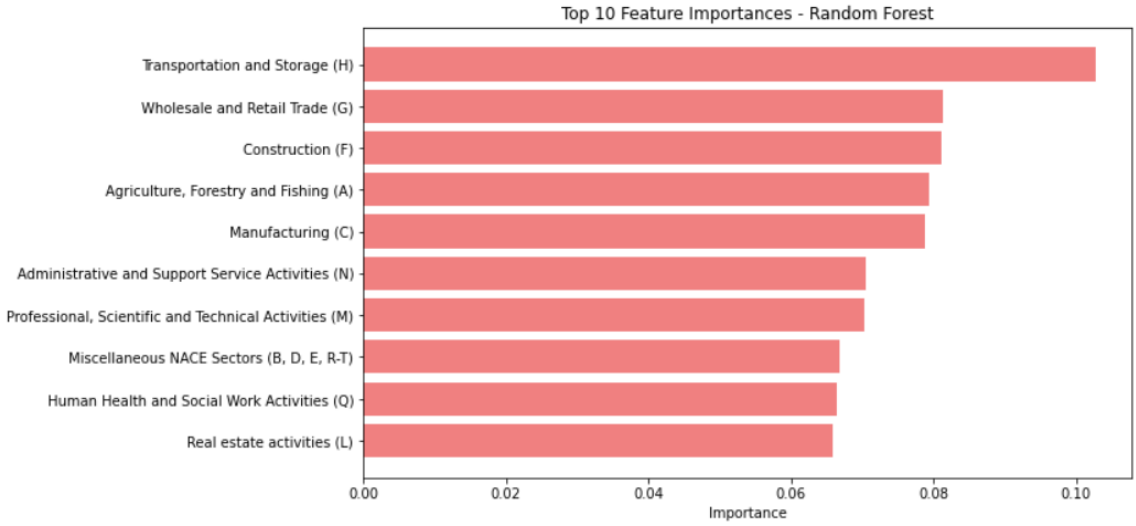

Figure 5: Feature Importance of Random Forest Model

Feature importance analysis provided insights into the factors influencing refugee employment in this sector. Figure 5 shows the top 10 feature importance.

- Transportation and Storage (H): It is most influential factor with an importance score of 0.102648. This indicates that employment trends in transportation might be closely connected to those in the Accommodation and Food Service sector. This correlation

is probably due to the dependency of the sector on logistics and mobility, both for supplies and customers.

- Wholesale and Retail Trade (G): This achieved an importance score of 0.08134. This may suggest that changes in retail and wholesale demand influence refugee employment opportunities in the accommodation and food services sector, possibly due to similar workforce skills or shared economic factors.
- Construction (F) and Agriculture, Forestry and Fishing (A): It showed significant importance scores of 0.081166 and 0.07939 respectively. These sectors could contribute seasonal or skill specific employment opportunities for refugees which is aligning with demand patterns in Accommodation and Food Services.
- Manufacturing (C) and Administrative and Support Service Activities (N): Achieved an importance scores of 0.078 and 0.07 respectively. It highlighted the significant role of Industrial and support services in shaping employment trends.
- Professional, Scientific and Technical Activities (M), Miscellaneous NACE Sectors (B, D, E, R-T), Human Health and Social Work Activities (Q), and Real Estate Activities (L): Their importance scores range is between 0.07 to 0.065. It showed that diverse set of industries that indirectly influence employment in Accommodation and Food Service sector.

This insight suggests more general NACE indicators that influence refugee employment in Accommodation and Food Service sector. This feature importance hierarchy shows that refugee employment in the Accommodation and Food Service sector is most strongly influenced by sectors with direct operational connections (transportation, wholesale) and those reflecting general NACE activity (construction, manufacturing) which provide insights for targeted employment support strategies.

6.2 Gradient Boosting Model

Gradient Boosting Model was implemented to predict the employment levels of the Ukrainian refugees in the Accommodation and Food Service sector of Ireland. In the first round of training itself, the model gives perfect accuracy at 100% which is R-squared score of the training is 1.0000 with RMSE 0.1129 and MAE 0.0908. This exceptionally high accuracy was a clear indicator of overfitting, as the model was very closely tailored to the training data, which negatively impacted its generalization capacity on test data. On the test set, the model significantly performed low as gave a R-squared of 0.9023, RMSE of 1666.6230, and MAE of 704.9361. This had to be dealt with by hyperparameter tuning. Initial hyperparameter tuning addressed this issue by optimizing learning rate (0.1), maximum depth (3), and number of estimators (100). After this tuning, the Gradient Boosting model was able to achieve test scores of R-squared 0.9867 which is 98.67% accuracy, RMSE 393.2624, and MAE of 332.7844. It closed in on its initial high accuracy at a controlled level of overfitting. Such improved model performance would rather suggest that Gradient Boosting is an excellent method for predicting refugee employment levels when optimizing toward generalized performance.

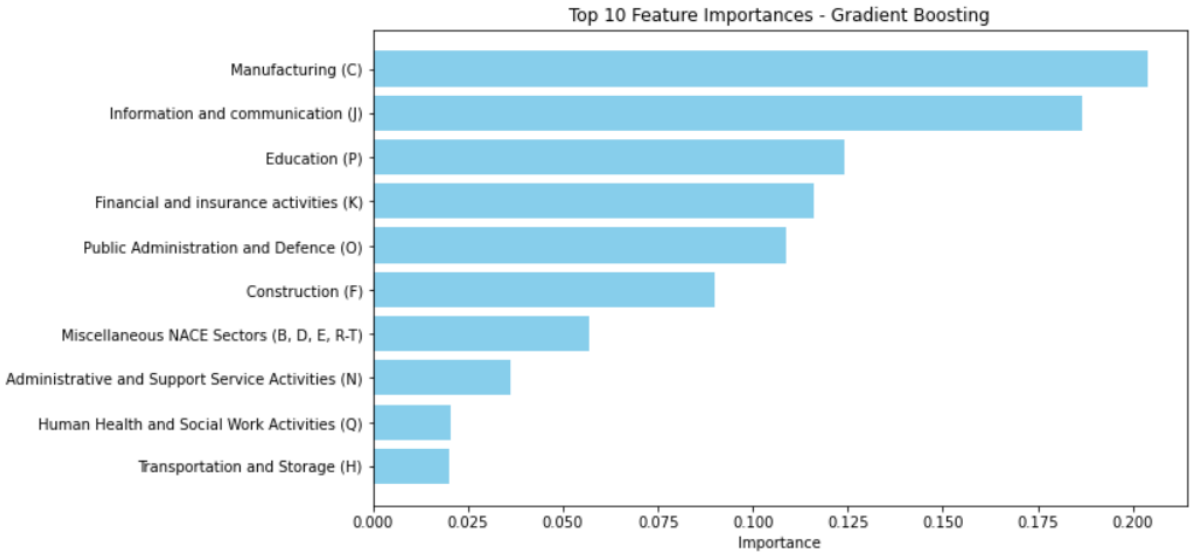

Figure 6: Feature Importance of Gradient Boosting

The feature importance analysis provides additional insights into the sectors most closely related to employment levels in the Accommodation and Food Service sector.

- Manufacturing (C): It is most influential sector with an importance score of 0.204069. This indicates that employment trends in manufacturing have a substantial impact on Accommodation and Food Service sector.
- Information and Communication (J): This achieved an importance score of 0.186781. It shows that employment shifts in Information and Communication influenced the Accommodation and Food Service sector. It suggests strong linkages between digital sector growth and the employment patterns in Accommodation and Food Service.
- Education (P) and Financial and Insurance Activities (K): Importance scores are 0.123934 and 0.116028 respectively. Showed significant influence, this suggest that labour demand in education and finance affects refugee employment.
- Public Administration and Defence (O) and Construction (F): Their importance scores are 0.108722 and 0.089981 respectively. Government programs and initiatives aimed at refugee integration have a measurable impact on employment outcomes in Accommodation and Food Services for Ukrainian refugees. Many refugees find employment in Construction, which provides them with entry-level work that enables them to establish a foothold in the job market, eventually contributing to employment in Accommodation and Food Services.
- Miscellaneous NACE Sectors (B, D, E, R-T), Administrative and Support Service Activities (N), Human Health and Social Work Activities (Q), Transportation and Storage (H): Importance scores are 0.056742, 0.036062, 0.020518 and 0.019942 respectively. These are the moderately influential sectors collectively create a supportive ecosystem for Ukrainian refugees in Ireland's Accommodation and Food Service sector.

This feature importance analysis shows that the employment of refugees in the Accommodation and Food Service sector is most influenced by manufacturing and Information

and Communication (digital) sectors, followed by educational and financial sectors, thus providing different insights than that obtained from the Random Forest model analysis.

6.3 Neural Network Model

The Neural Network model initially struggled to predict employment levels for Ukrainian refugees in Ireland's Accommodation and Food Service sector with a poor fit: the training R-squared was only 0.0530, with an high RMSE of 3386.8845; the model was even worse with a test R-squared of -0.3639, confirming severe underfitting with RMSE 3985.5260, MAE 3658.6542. After hyperparameter tuning, using the best parameters: tanh activation function, hidden layers of size (50,50), constant learning rate, 1000 iterations, and sgd solver, the model improved, and performances scored a cross-validation R-squared score of 0.8314 and a final test R^2 of 0.6673 which is only 66.73% accuracy. While this tuning increased accuracy the Neural Network still underperformed relative to the Random Forest and Gradient Boosting models which could be an indication that it might require a larger dataset to fully capture the complex patterns in employment data for that sector. Unlike Random Forest and Gradient Boosting which provide feature importance for predicting employment levels of Ukrainian refugees in Accommodation and Food Service sector, the Neural Network which implemented by using MLPRegressor that does not inherently provide importance metrics.

6.4 Model Comparison

In this study specifically three models (Random Forest, Gradient Boosting and Neural Networks) were used to compare their predictive accuracy using R-squared, RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) and Learning curves.

Table 2: Comparison of Models

Model	Accuracy
Random Forest	98.29%
Gradient Boosting	98.67%
Neural Network	66.73%

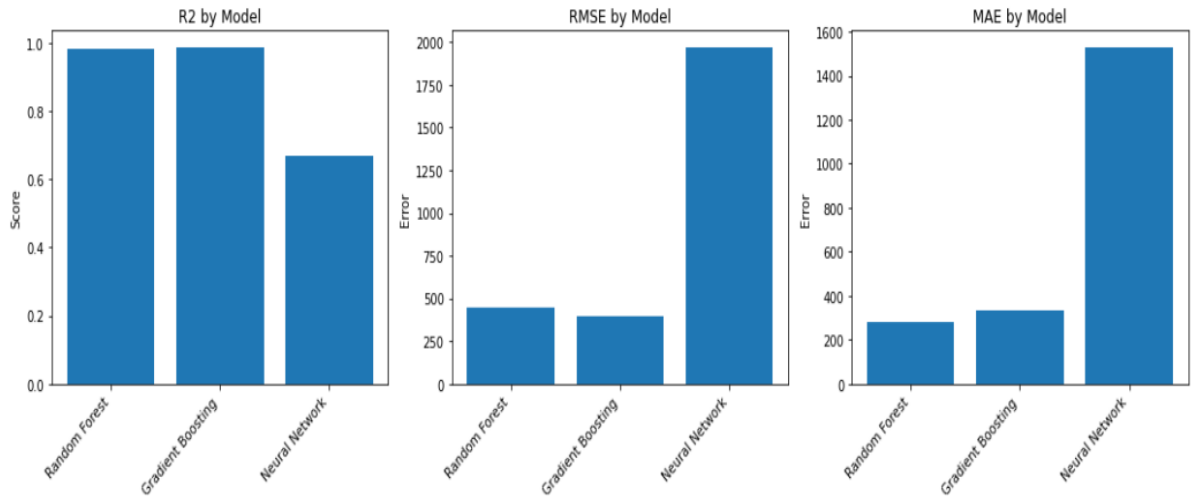

Figure 7: Comparison Analysis of Models

Table 2 showed that Model and their accuracy, Gradient boosting has a highest accuracy than Random Forest and Neural Networks. In figure 7, R-squared chart showed that Gradient Boosting achieved highest accuracy at 98.67% closely followed by Random Forest at 98.29%. Both models captured the variance in employment levels effectively and demonstrated strong predictive accuracy. Whereas in Neural Network lagged significantly with R-squared 0.6673 which is 66.73% accuracy. Indicated that it struggled to explain the variation in the data. RMSE by model chart showed that Gradient Boosting had the lowest RMSE score which is 393.2624 followed by Random Forest with RMSE value of 446.6. These low RMSE values shows that both models produced predictions close to the actual values on average. However, in Neural Network RMSE value is 1968.2931 showing larger prediction errors and weaker performance. Also, in MAE by model chart showed that Random Forest has the lowest Mean Absolute Error value which is 283.03. Gradient Boosting has the MAE value 332.7844 which is slightly more than Random Forest. However, in Neural Network shows the higher absolute prediction errors which is 1527.9727 values of MAE. Confirmed its lower accuracy in predicting refugee employment levels. Because there was a discrepancy in the performance metrics, with Random Forest having the lowest MAE of 283.0390, while Gradient Boosting had a better R-squared of 0.9867 and RMSE of 393.2624, learning curve analysis should be performed to identify the most reliable model for the prediction of Ukrainian refugee employment in the Accommodation and Food Service sector.

6.4.1 Learning Curve

In Learning Curve, the red line shows the training score of model's performance on training data. Green line shows cross-validation of model's performance on unseen data and Convergence of red line and green line indicated that good balance between bias and variance.

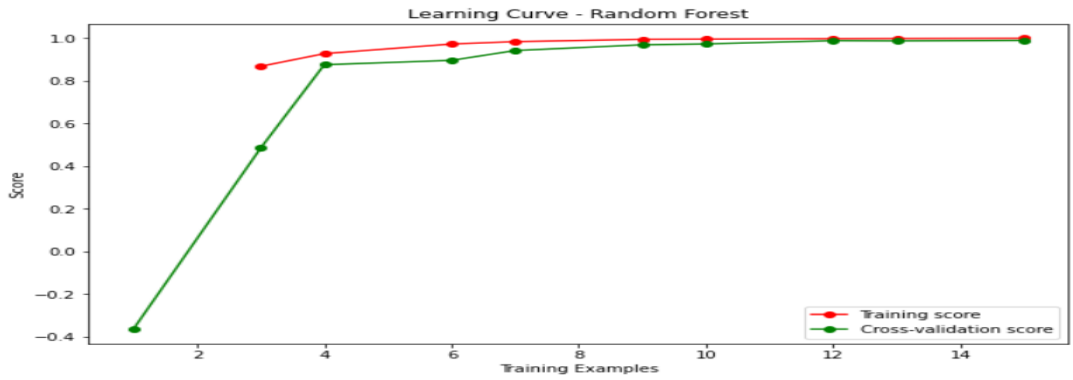

Figure 8: Learning Curves of Random Forest

Figure 8 Random Forest learning curve shows that both training and cross-validation scores towards 1 with the increase in training data. Training score (red line) starts at 0.85 and it gradually increased to near 1. Green line which is Cross-validation score improved gradually from -0.4 to 0.95. The line converge at approximately 0.98. This indicates high accuracy of Random Forest and maintain stable generalization, capturing patterns in the data without overfitting. This suggest that by using the available data to predict employment levels for Ukrainian refugees in Ireland's accommodation and food service sector.

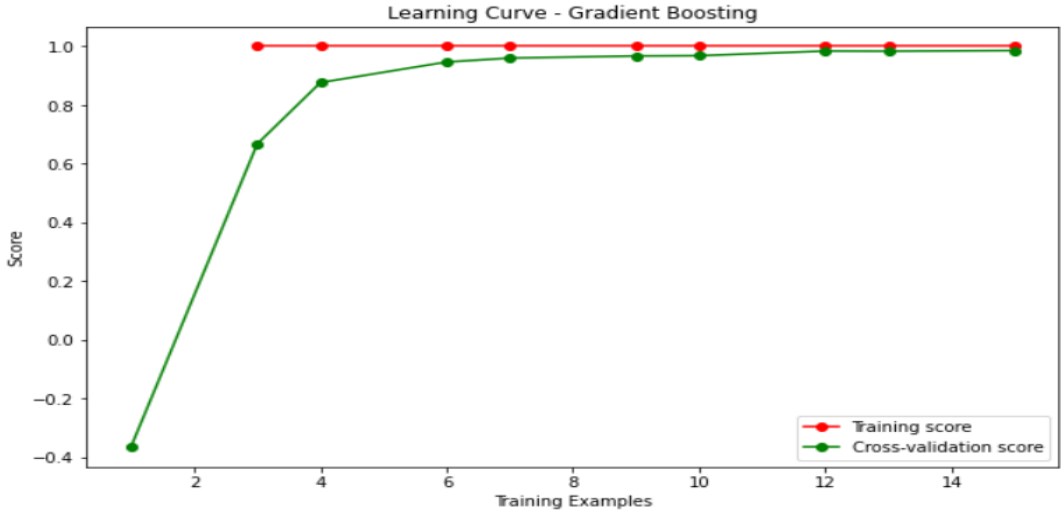

Figure 9: Learning Curves of Gradient Boosting

Figure 9 Gradient Boosting learning curve shows the strong performance with training and cross validation scores converging near 1. Red line which is training score maintained high performance around 1.0. Cross-validation (green line) improved from -0.35 to 0.98. Line converge is at approximately 0.99. Minimal gap between final training and cross-validation scores indicated that optimal generalisation. This shows that Gradient boosting accurately captured complex patterns in the refugee employment data. The model can minimize the error at each iteration.

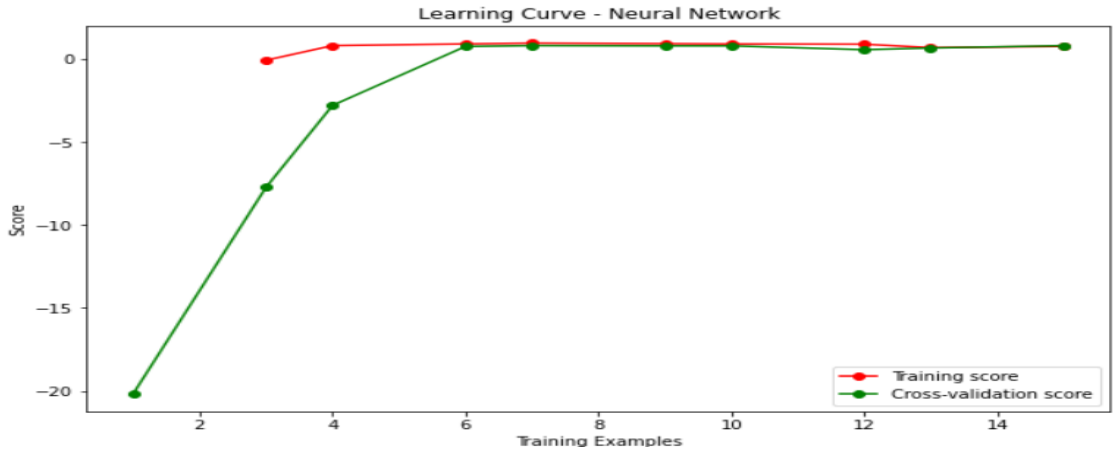

Figure 10: Learning Curves of Neural Network

Figure 10 shows Neural Networks learning curve, it shows that a high training score but lower cross validation score. Red line stabilizes around 0.5, Green line showed poor initial performance at -20 and improved by marginally only. This suggests an overfitting, that model performed well on training data but failed to generalize to a new data.

Despite Random Forest showing better Mean Absolute Error (MAE) of 283.0300 compared to Gradient Boosting's MAE of 332.7844. Gradient Boosting model was selected as the optimal model, slightly outperforming Random Forest and significantly outperforming Neural Network's poor performance which was R-squared = 0.6673, Root Mean Squared Error = 1968.2931 and Mean Absolute Error = 1527.9727. This selection was justified by Gradient Boosting's marginally higher R-squared value of 0.9867 (98.67% accuracy) compared to the one achieved by Random Forest's R-squared value: 0.9829 (98.29% accuracy) and with a lower Root Mean Squared Error (RMSE) of 393.2624 compared to Random Forest with 446.6000. The Learning Curves analysis supported this selection by demonstrating a more stable convergence of Gradient Boosting and tighter a gap between training and cross-validation scores, which indicated a better generalization and stability in predicting the employment levels of Ukrainian refugees in Ireland's Accommodation and Food Service sector. The slight advantage in R-squared, 0.0038 higher, and in RMSE, 53.3376 lower, with more stable convergence between training and cross-validation scores in learning curves, made Gradient Boosting slightly more reliable for this specific prediction task, while both are substantially better than the approach using the Neural Network.

6.5 Discussion

The study of predicting Ukrainian refugee employment levels in Ireland's Accommodation and Food Service sector utilized R-squared, RMSE, MAE, and Learning Curves to evaluate model performance. R-squared measured how effectively each model explained employment variations using other NACE sectors employment levels as predictors. The R-squared value of 0.9867 derived from the Gradient Boosting model proves that 98.67% of employment changes within the Accommodation and Food Service sector are explained by the levels of employment in other NACE sectors. The RMSE gives the magnitude of the prediction error. The smaller RMSE of Gradient Boosting, 393.2624 as compared to the 446.6000 from Random Forest, points out that this model provides more accurate predictions of refugee employment numbers.

MAE calculated the average prediction error of the models, and on that count too, Random Forest had performed better in predicting the absolute number of employments, with an MAE value of 283.0300 against Gradient Boosting, which had 332.7844. Learning curves for the 2022-2024 analysis period demonstrated how each model performed with increasing employment data, with Gradient Boosting showing tighter convergence between training and cross-validation scores near 0.99, indicating better generalization in predicting refugee employment levels.

Although, Random Forest showed better MAE, Gradient Boosting's higher R-squared value (0.0038 higher) and lower RMSE (53.3376 lower), combined with more consistent convergence between training and cross-validation scores in learning curves, established it as marginally more reliable for predicting refugee employment levels in the Accommodation and Food Service sector. Given Gradient Boosting's superior predictive performance and more reliable generalisation, its feature importance analysis was selected to identify influential sectors. This analysis quantified sector influence, revealing Manufacturing (0.204069), Information and Communication (0.186781), and Education (0.123934) as the key sectors impacting refugee employment levels in the Accommodation and Food Service sector. These feature importance scores provided specific, quantitative measures of each sector's contribution to predicting refugee employment levels, with Manufacturing showing the strongest influence at 20.4%, followed by Information and Communication at 18.7%, and Education at 12.4%.

7 Conclusion and Future Work

7.1 Conclusion

This study has successfully achieved its aim of predicting employment levels of Ukrainian refugees in Ireland's Accommodation and Food Service sector between 2022 to 2024 using machine learning methods, leveraging employment trends from other NACE sectors as predictors. In addressing the research question of which model achieves highest prediction accuracy, the comparison of performance metrics revealed Gradient Boosting as the superior model, slightly outperforming Random Forest and significantly outperforming Neural Network. Gradient Boosting demonstrated the highest R-squared (0.9867) and lowest RMSE (393.2624). While Random Forest showed slightly better MAE (283.0390 compared to Gradient Boosting's 332.7844), the marginal advantage in R-squared (0.0038 higher) and RMSE (53.3376 lower) by Gradient Boosting, combined with learning curve analysis, confirmed its slight superiority through optimal convergence between training and cross-validation scores, with both lines stabilizing near 0.99. This indicated better generalization capabilities and more reliable predictions compared to Random Forest's persistent gap between training and validation performance. Neural Network significantly underperformed with an R-squared of 0.6673, RMSE of 1968.2931, and MAE of 1527.9727, proving substantially less effective for this prediction task, with its performance notably inferior to both Gradient Boosting and Random Forest models.

The second objective of identifying influential NACE sectors was accomplished through Gradient Boosting's feature importance analysis, which revealed Manufacturing (Importance Score= 0.204069), Information and Communication (Importance Score=0.186781), and Education (Importance Score=0.123934) as the sectors most significantly impacting employment levels in the Accommodation and Food Service sector. These findings highlight critical sector relationships:- manufacturing employment trends strongly influence support

service demand, the Information and Communication sector's impact reflects the growing importance of digital capabilities, and the Education sector's influence emphasizes the role of skill development in employment opportunities.

Feature importance scores from Gradient Boosting quantified the relative influence of different NACE sectors on the level of employment in the Accommodation and Food Service sector. The strongest influence was Manufacturing at 0.204069, followed by Information and Communication 0.186781, then Education 0.123934 - both showing strong predictive relationships with the level of employment in the Accommodation and Food Service sector.

7.2 Future Work

Future work could address specific limitations identified in this study, particularly focusing on improving the poor performance of Neural Network (R-squared = 0.6673) in predicting Ukrainian refugee employment levels. The dataset could be expanded beyond 2022-2024 to include additional employment indicators from the identified influential sectors (Manufacturing, Information and Communication, and Education), more detailed sector-specific employment characteristics, and longer time series data. Model improvements could focus on addressing underperformance of Neural Network through expanded training data and testing advanced architectures specifically designed for employment time series prediction, while exploring ensemble methods combining the strengths of Gradient Boosting (R-squared = 0.9867) and Random Forest (R-squared = 0.9829). Feature engineering could be enhanced by developing additional features from the three most influential sectors identified by Gradient Boosting's feature importance analysis:- Manufacturing (0.204069), Information and Communication (0.186781), and Education (0.123934), along with temporal features to capture seasonal employment variations. Real-time prediction capabilities could be implemented through continuous model updating with new employment data and automated updates for employment level predictions in the Accommodation and Food Service sector. These enhancements would build upon the current study's identification of key influential sectors and model performance metrics to potentially improve prediction accuracy for Ukrainian refugee employment levels.

References

Abramov, M. (2024). *Maximizing Performance: AI Model Optimization Techniques*.

Ahani, N. (2023). *Dynamic Placement in Refugee Resettlement*.

Ahmad, H. S. A. I. (2023). *A COMPARATIVE ANALYSIS OF GRADIENT BOOSTING, RANDOM FOREST AND DEEP NEURAL NETWORKS IN INTRUSION DETECTION SYSTEM*.

Aiken, M. (2020). *A neural network t al network to predict civilian unemplo edict civilian unemployment r yment rates*.

Auret, L., & Aldrich, C. (2012). Interpretation of nonlinear relationships between process variables by use of random forests. *Minerals Engineering*, *35*, 27–42. https://doi.org/10.1016/j.mineng.2012.05.008

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2019). *A Comparative Analysis of XGBoost*. https://doi.org/10.48550/ARXIV.1911.01914

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*(3), 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Central Statistics Office Ireland. (2024). Ukrainian Population in Ireland Report. *Central Statistics Office Ireland*. https://www.cso.ie/reports/ukraine2024

Davidescu, A. A., Apostu, S.-A., Paul, A., & Casuneanu, I. (2020). Work Flexibility, Job Satisfaction, and Job Performance among Romanian Employees—Implications for Sustainable Human Resource Management. *Sustainability*, *12*(15), 6086. https://doi.org/10.3390/su12156086

Department of Enterprise, Trade and Employment. (2024). Labour Market Integration Report. *Government of Ireland*.

ESRI, McGinnity, F., Enright, S., ESRI, Quinn, E., ESRI, Maitre, B., ESRI, Privalko, I., ESRI, Darmody, M., ESRI, Polakowski, M., & ESRI. (2020). *Monitoring report on integration 2020*. ESRI. https://doi.org/10.26504/bkmnext403

European Commission. (2024). Refugee Integration Outcomes Study. *Office of the European Union*. https://doi.org/10.2767/123456

Groeger, A. (2023). *Forecasting Bilateral Refugee Flows with High-dimensional Data and Machine Learning Techniques*. https://www.itflows.eu/wp-content/uploads/2023/03/1387.pdf

Hadi, sahal. (2023). *P R E D I C T I N G E C O N O M I C S E L F - S U F F I C I E N C Y A M O N G R E F U G E E S*. https://arno.uvt.nl/show.cgi?fid=171234

Huo, X., Hao, K., Chen, L., Tang, X., Wang, T., & Cai, X. (2022). A dynamic soft sensor of industrial fuzzy time series with propositional linear temporal logic. *Expert Systems with Applications*, *201*, 117176. https://doi.org/10.1016/j.eswa.2022.117176

Johnson, A. (2020). *Predicting Housing Stability Among Homeless Individuals Using Machine Learning Techniques*. Journal of Social Services Research.

kiwk, md. (2024). *HyperParameter Tuning: Fixing Overfitting in Neural Networks*.

Liu, A. J., Mukherjee, A., Hu, L., Chen, J., & Nair, V. N. (2022). *Performance and Interpretability Comparisons of Supervised Machine Learning Algorithms: An Empirical Study* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2204.12868

McGuinness, S., K., E. (n.d.). Sectoral Employment Patterns Among Refugees in Ireland. *Economic and Social Research Institute*. https://www.esri.ie/publications/789

McQuinn, K. (2024). *QUARTERLY ECONOMIC COMMENTARY*. MACRO ECONOMIC FORECASTING.

Nalevanková, P., Fleischer, P., Mukarram, M., Sitková, Z., & Střelcová, K. (2023). Comparative Assessment of Sap Flow Modeling Techniques in European Beech Trees:

Can Linear Models Compete with Random Forest, Extreme Gradient Boosting, and Neural Networks? *Water*, *15*(14), 2525. https://doi.org/10.3390/w15142525

Patel, A. (2023). *Healthcare access prediction for refugees*. https://doi.org/10.1016/j.jbi.2023.104097

Schmitt, M. (2022). *Deep Learning vs. Gradient Boosting: Benchmarking state-of-the-art machine learning algorithms for credit scoring* (No. arXiv:2205.10535). arXiv. http://arxiv.org/abs/2205.10535

Teytelboym, A. (2017). *Placement Optimization in Refugee Resettlement*.

UNHCR. (2023). *Global Trends report 2023*. https://www.unhcr.org/global-trends-report-2023