Understanding the Changes in the Wages Of UK Employees

A dissertation submitted In partial fulfilment
of the requirements for the degree of

Master of Science Artificial Intelligence and Data Science
in the University of Hull
By

Olivia C. Okoro

December, 2023

Understanding the changes in wages of UK employees

Abstract

Accurate prediction of annual wages holds paramount importance in economic forecasting and policy formulation. This study conducts a thorough analysis of regression models for multi-industry annual wage prediction. A range of models: Linear Regression, Random Forest, Decision Tree, Ridge Regression, Support Vector Regression with diverse kernels, XGBoost, and a Feedforward Neural Network, undergo rigorous evaluation and optimization to determine the most effective model. Methodologies encompass preprocessing, feature selection, model training, hyperparameter tuning, and cross-validation for robustness assessment. Key performance metrics, namely RMSE, MAE, and R2, guide model comparison. Results indicate distinct model performances. Specifically, after optimization, the Random Forest Regression emerges as the most accurate, yielding an RMSE of £4540.76, MAE of £2942.13, and an R2 of 0.6112. This underscores its efficacy in annual wage prediction. The study further broadens its understanding by providing future forecasts for annual wages from 2023 to 2027 using the Random Forest model: 2023: £28021.72, 2024: £28631.49, 2025: £29241.27, 2026: £29851.04, 2027: £30460.81. Conversely, the Support Vector Regression using the Sigmoid kernel demonstrates restricted predictive capabilities, underscoring the integral role of model selection. This research compares wage prediction models, favouring Random Forest Regression. It's a useful reference for accurate predictions in labour economics.

Introduction/Background

Understanding of employee wage dynamics in the United Kingdom is very crucial for policymakers, economists, and stakeholders tasked with shaping societal and economic conditions (Koeniger et al., 2007). These parties have rigorously examined this discourse, delving into wage trends, the intricate factors that cause fluctuations, disparities observed across various industries and demographic groups, the impact on economic conditions and policies, and the development of predictive models for anticipating wage changes (Chen et al., 2019).

Compositional effects are observed in the workforce when there are shifts in employment, such as people transitioning between jobs or joining and leaving the workforce. These changes often impact the overall characteristics of those employed, influencing variables like gender distribution, tenure, and educational levels. The composition of the workforce which can be in terms of individual preferences or job demands can largely affect wages trend. Investigations into the role of composition effects in shaping average wage growth within a singular nation, exemplified by Abel et al. (2016), indicate that these effects yielded a marginally negative impact on wage growth in the UK during 2014. This impact persisted, though diminishingly, into 2015. In 2019 Kiss et al, showed that the impact of composition effects resulting from shifts in age, gender, or educational qualifications within the labour force, considered largely independent of economic changes, had a limited influence on observed wage dynamics. Also, the study reveals that the influence arising from work-related attributes, mostly influenced by economic-related shifts such as evolutions in workforce categories, holds increased significance in agreement with a publication by the Bank of England (2016) discussing the changing composition of the UK workforce and its impact on wages. In 2017 Bhattarai examined the UK labour market, revealing how psychological profiles influence labour participation decisions, while local market conditions strongly impact employment prospects. He explained 92% of wage rate variance showcasing understanding of the factors affecting salaries and offers unbiased labour supply elasticity estimates based on the British Household Panel Survey (BHPS) data. Also, in 2020 Dolton et al examined the relationship between public and private sector wages in the UK. The researchers highlighted how public sector wages adjust to align with private sector wages over time and notes short term effects where public sector decisions influence private sector wages. Understanding these wage dynamics is crucial for policymakers as it helps them anticipate potential ripple effects and make informed decisions that consider the broader economic implications.

More recent research by De Almeida Vilares & Reis (2022) examined the impact of sectoral bargaining on wage distribution over two decades. This revealed a noteworthy trend: a rise in worker bargaining power at lower skill levels while declining at mid and high levels. This shift has compressed wage distributions, highlighting the growing importance of sectoral bargaining in wage determination even during economic downturns like COVID-19 pandemic (Mayhew & Anand, 2020, HM Treasury, 2020). Understanding these dynamics sheds light on how bargaining power influences wage distribution and productivity in diverse economic contexts.

This existing research while providing insightful perspectives, leaves certain critical knowledge gaps. For instance, Kiss et al. (2019) and Abel et al. (2016), provide insights into the limited impact of demographic shifts on observed wage dynamics. However, further exploration into the intricate relationships between these shifts and job-related characteristics, notably influenced by economic changes would be of importance. While Bhattarai's 2017 study illuminates psychological and local market influences on wage rates and employment prospects, a deeper investigation into their interplay with broader economic trends is necessary. Also, the recent research by de Almeida Vilares & Reis (2022) unveils trends in sectoral bargaining's influence on wage distribution, particularly in altering worker bargaining power across skill levels. Yet, an in-depth understanding of how economic conditions shifts in bargaining power and their prolonged impact on sector-specific wage distributions remains shallow. Addressing these gaps would foster a more holistic understanding of the intricate forces shaping wage dynamics among UK employees. Our research aims to bridge the existing knowledge gaps by analysing different factors affecting wage dynamics among UK employees. While prior studies have offered insights into individual aspects of wage determinants, our approach synthesizes these insights to create a holistic model. By considering demographic shifts and job-related characteristics on wage rates, our research aims to create a comprehensive framework.

These gaps will be addressed by providing insights pythonthe following questions:

1. What are the historical patterns or trends in wage changes among diverse industries and demographic groups in the UK?
2. How can traditional regression models and a feedforward neural network enhance the prediction of wage fluctuations based on historical data?
3. To what extent do economic indicators, labour market conditions, and policy interventions influence wage changes?

We seek to extensively understand the various impacts of economic indicators, labour market conditions, and policy interventions on wage dynamics among UK employees. Our research methodology includes preprocessing, feature selection, model training, hyperparameter tuning, and cross-validation, ensuring robustness in our analysis. Through predictive modelling, we unravel the intricate interplay between these variables and broader economic trends. Our study aims to contribute a holistic understanding of how these collective factors shape wage trends, offering insights crucial for policymakers, economists, and stakeholders in shaping societal and economic conditions.

Methodology

Data Collection and Preprocessing

The datasets used in this analysis were wages data 2018-2022 and Westminster Parliamentary Constituencies 2021 data. These were obtained from NOMIS - Office for National Statistics (ONS) detail annual wages of UK employees spanning from 2018 to 2022. Using Python and Jupyter notebook, rigorous data cleaning techniques were applied to ensure data integrity. This involved handling missing values, ensuring uniform data formats, removing duplicates, and eliminating irrelevant or redundant entries. Imputation strategies or deletion techniques were employed where appropriate. Features such as ‘Hourly pay gross’, ‘Hours worked total’, ‘sex’, ‘year,’ ‘work type’, and various 'levels of qualifications' were meticulously chosen based on their perceived influence on predicting 'annual pay gross.' Categorical variables were encoded, numeric features were scaled or normalized, and potential outliers were addressed through robust statistical techniques. These transformations aimed to prepare the dataset for accurate model training and prediction.

The Westminster Parliamentary Constituencies 2021 data comprises information about parliamentary constituencies in the UK, detailing their names, unique identifiers, geographic coordinates (longitude and latitude), area sizes, and geometric properties. It allows for geographical analysis, including mapping and spatial assessments, to understand the distribution and properties of these constituencies while the wages data 2018-2022 contain information related to different parliamentary constituencies in the UK, including various attributes such as hourly pay, total hours worked, annual gross pay, gender (sex), year, work type, educational qualifications, occupation categories, and ethnicity. It seems to provide a comprehensive view of the demographics and employment-related statistics within these constituencies, allowing for detailed analyses of factors influencing wages, work patterns, educational backgrounds, occupational distributions, and diversity in the workforce across different regions or constituencies.

Statistical Analysis

The pre-processed numeric datasets were used to carry out statistical analysis using Pythons. The describe() function enabled the determination of min, max, and count values, along with the mean for each of the two datasets. Both datasets were then merged in conjunction with a common identifier to create a new dataset (merged_df). This information was visualised for each set of comparison like average wage per male/female, annual wage versus year using Seaborn's boxplot, calculating the correlation matrix for the selected features and creating scatter plots between two numerical features. Features containing outliers were determined using plots and functions.

'outliers_treatment(df.col)' and 'treat_outliers(Numeric_wages_df, col_list)' were created to treat outliers.

The heatmap in Figure 1 portrays correlations between 'Annual gross pay' and related variables. It highlights strong positive correlations with 'Hourly pay gross (0.70)', 'Hours worked total (0.59)', 'work type (0.53)', and 'sex (0.40)', indicating their direct influence on higher wages. Conversely, weaker correlations exist with educational levels, suggesting a less direct impact on pay. Interestingly, an inverse relationship ('anticorrelation') is observed with 'ethnicity/race', implying a potential association with lower annual pay. Overall, the heatmap illustrates key factors affecting wages, emphasizing the dominant roles of hourly pay, hours worked, work type and gender.

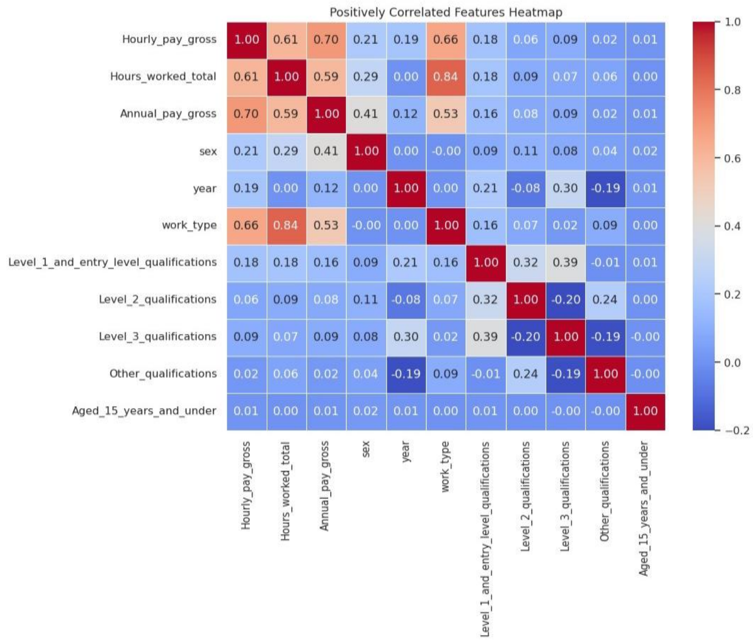

Figure 1:Heatmap of the correlation matrix of the features 'Hourly_pay_gross' showing positive correlation of 0.70

Model Building

We describe the regression models utilized for predicting future annual wages and outline the strategies applied for hyperparameter tuning and model evaluation. The regression models employed encompassed Linear Regression, Support Vector Regression (SVR), Random Forest Regression, Ridge Regression, XGBoost Regression, and a feedforward neural network (deep learning model). To enhance the models' capabilities, we carefully optimized their performance through hyperparameter tuning. For instance, Decision Tree Regressor, Ridge Regression, and Support Vector Regression underwent grid searches for parameter optimization, exploring various configurations for parameters like maximum depth, minimum samples split, and leaf. Additionally, the Support Vector Regression utilized kernels such as poly, rbf, and sigmoid to explore different function transformations. For the Random Forest Regressor, we employed RandomizedSearchCV, systematically testing multiple hyperparameter combinations like the number of estimators, maximum depth, and minimum samples split. This exploration aimed to identify the most effective parameter sets, enhancing the models' predictive power and precision in forecasting wage dynamics. Evaluation metrics including Mean Absolute Error, Root Mean Squared Error (RMSE), (MAE), and R-squared (R2) were used to assess and compare the performance of these models.

Results

Data Visualisation

Pointplots were instrumental in understanding the intricate relationships between different features and wage dynamics (Emerson et al, 2013). These plots, presenting pointwise relationships in the dataset, allowed for a clearer understanding of how various factors interact. In figure 2, we see notable trends between gender, work type, and annual wages. Specifically, the analysis revealed a consistent trend where men generally had higher annual wages compared to women over the observed period, indicating a gender-based wage disparity agreeing with study by Bowlus and Grogan (2009) on gender wage differential.

The yearly earnings of men showed a steady and gradual rise from 2019 to 2022, with a slight dip in 2020 attributed to the effects of COVID-19, followed by a subsequent sharp increase in 2021. In contrast, the annual wages of women displayed a distinct pattern with a sharp rise starting from 2019, a noticeable drop in 2020, an effect of COVID-19 and a subsequent spike in 2021. These observations highlight gender-based differences in wage trends and fluctuations over the analysed years agreeing with Cribb et al, (2022) in a 25 year study of income gap in Britain.

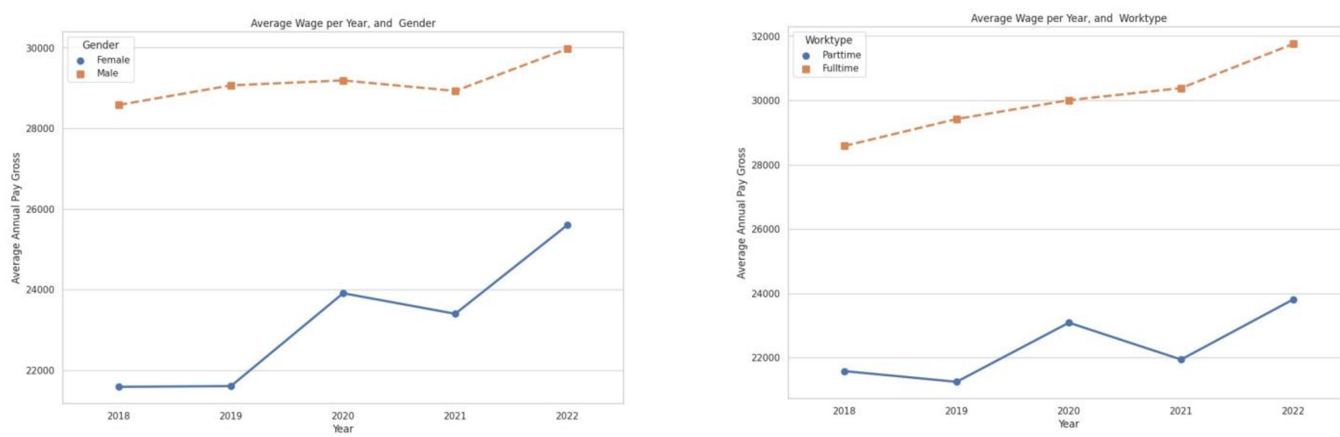

Figure 2: Pointplots of average wages per year vs gender and average wages per year vs work type.

Utilizing box plots provided a valuable means to visually comprehend wage distribution across diverse categorical variables. These plots effectively highlighted variations in wage distributions among different job categories, thereby shedding light on potential wage disparities prevalent across various occupations. This visual insight is imperative for addressing income inequality concerns and advocating for equitable wage practices within the workforce.

The median depicted in the box plot, averaging at £26,000 signifies that the majority of workers' wages are concentrated around this value. This finding notably from figure 3 and figure 4 align with a study from the Office of National Statistics (ONS), as reported by tastingbritain.co.uk, which stipulates that the median wage for the middle class stands at £26,800 (tastingbritain.co.uk, 2023).

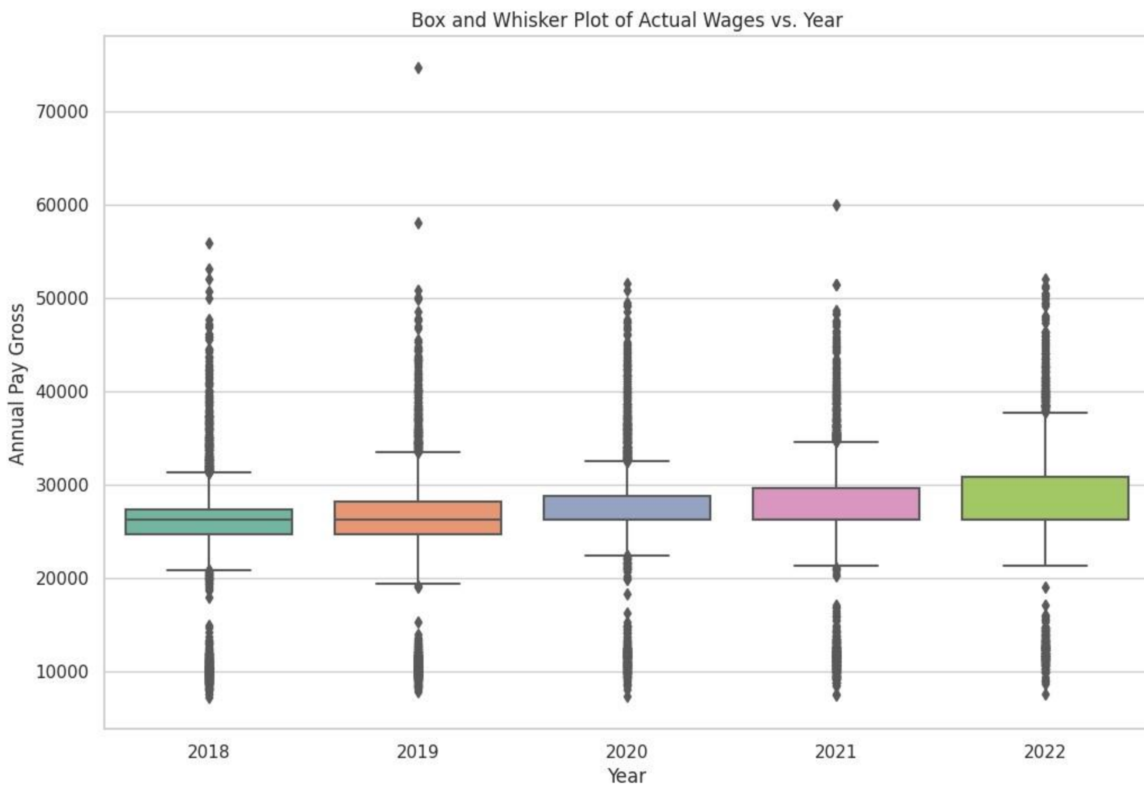

Figure 3: Box plots of annual wages vs year.

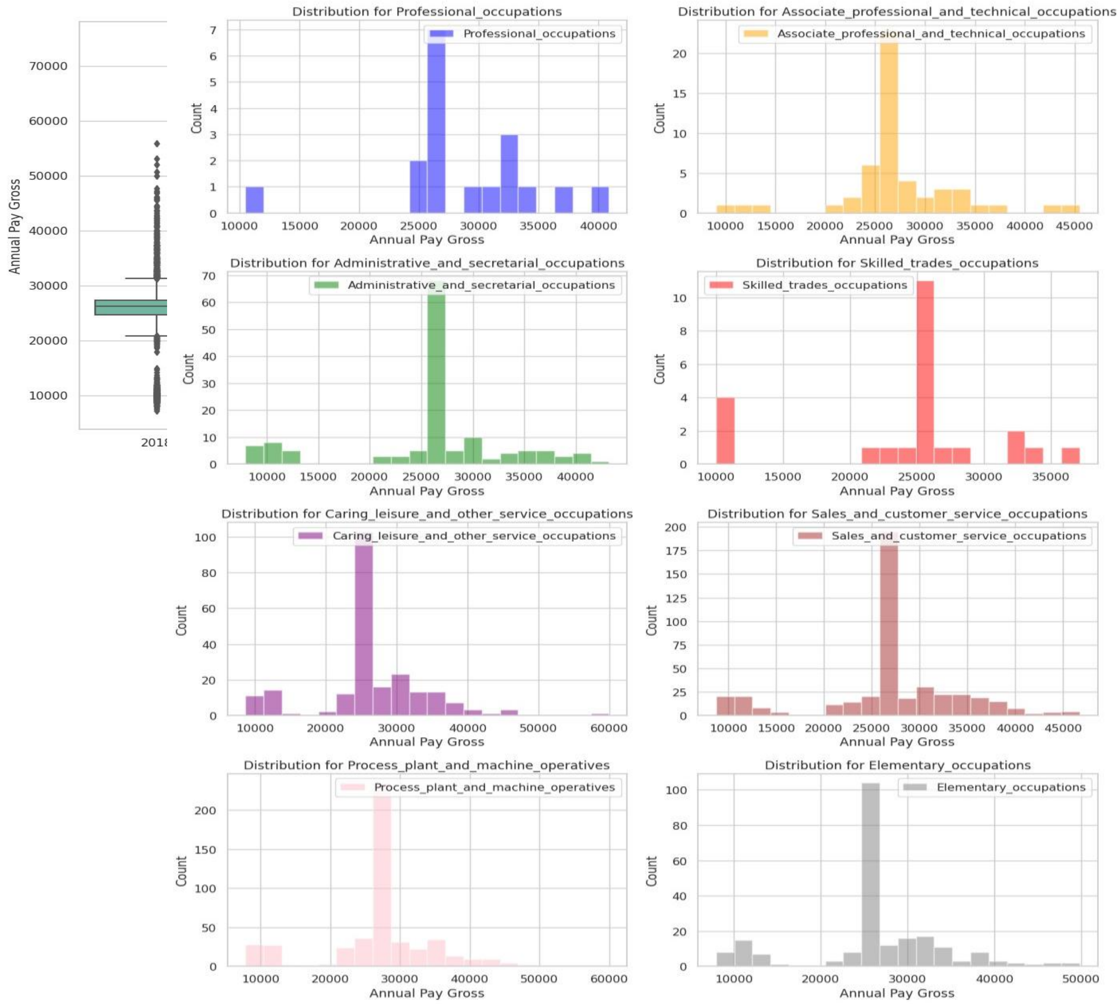

Figure 4: Histogram plot of occupations compared to annual pay gross.

The examination of annual wages across successive years in the UK presents a variety of wage dynamics reflective of economic shifts and labour market intricacies. The details are as follows:

Year 2018

The presence of outliers in this year indicates sporadic instances of both notably high and exceptionally low annual wages. The spread, as indicated by the whiskers extending up to 48000, emphasizing the considerable diversity in wage distributions.

Year 2019

Outliers persist, emphasizing instances of considerable wage disparities. The range of wages remains diverse, though slightly less spread out compared to the previous year, reflecting a narrowing distribution.

Year 2020

While outliers still exist, the distribution begins to show a discrete compressed spread. This suggests a potential convergence of wages, indicated by the lesser variation between lower and higher wage segments.

Year 2021

Continued outliers highlight persistent disparities in wages. However, the box plot indicates a slightly wider IQR, reflecting a re-expansion in wage variability despite the outliers.

Year 2022

The absence of outliers in this year suggests a more stabilized wage distribution. The limited variation, as demonstrated by the compressed spread, signifies a potentially narrowed wage spectrum compared to previous years.

Across these years, the data portrays a fluctuating landscape of wage distributions. The presence of outliers in various years signifies instances of extreme wage variations, while the compression or expansion of the interquartile range signifies periods of potential wage convergence or divergence.

Box plots comparing wages based on educational qualifications offered insights into how different levels of education correlate with salary disparities.

Understanding these variations is essential for devising strategies to bridge educational attainment-related wage gaps.

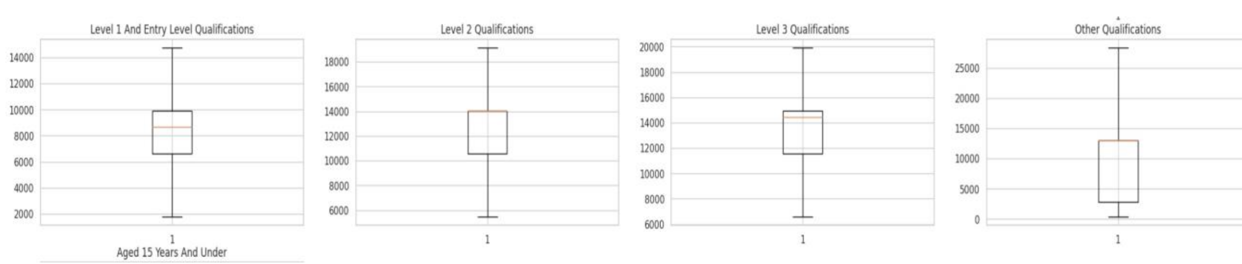

Figure 5: Box plots comparing wages based on qualifications.

These box plots (figure 5) showcase wage distributions across different educational qualifications. The concentration of wages within the IQR for each qualification level indicates a significant proportion of individuals earning within a specific income range. However, the presence of outliers, especially those at higher income levels, significantly widens the wage distributions, signifying substantial income disparities even within specific qualification categories. This observation underscores the influence of educational qualifications on wage levels, with outliers suggesting the potential for higher earnings, potentially influenced by additional factors beyond qualifications alone agreeing with study in 2018 by Alsulami revealing that higher educated individuals earn higher wages (Alsulami, 2018).

Scatter plot (figure 6), depicting the relationship between hourly gross pay and annual gross pay, provided relative insights into how working hours impact wages. It revealed the nature of the relationship between hourly pay and total hours worked. Patterns emerging from this plot could indicate the presence of hourly wage thresholds or the influence of overtime on overall annual pay with outliers on high end of wages and higher working hours showing that total hours worked leads to higher annual gross pay.

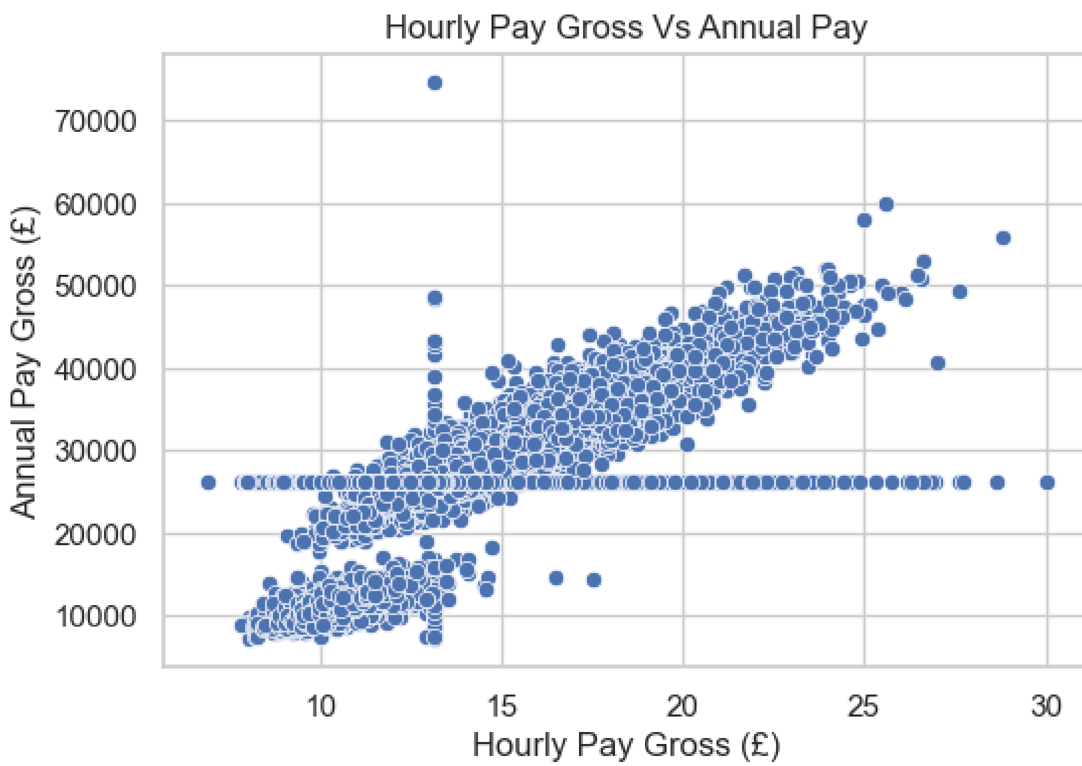

Figure 6: Scatter plot comparing annual pay gross and hourly pay gross.

Figure 7 depicts extreme variations in annual wages among parliamentary constituencies. Chelsea and Fulham exhibit high annual wages, exceeding £60,000, while Tottenham shows notably lower wages, below £8,000. These visual highlights significant income disparities across UK constituencies.

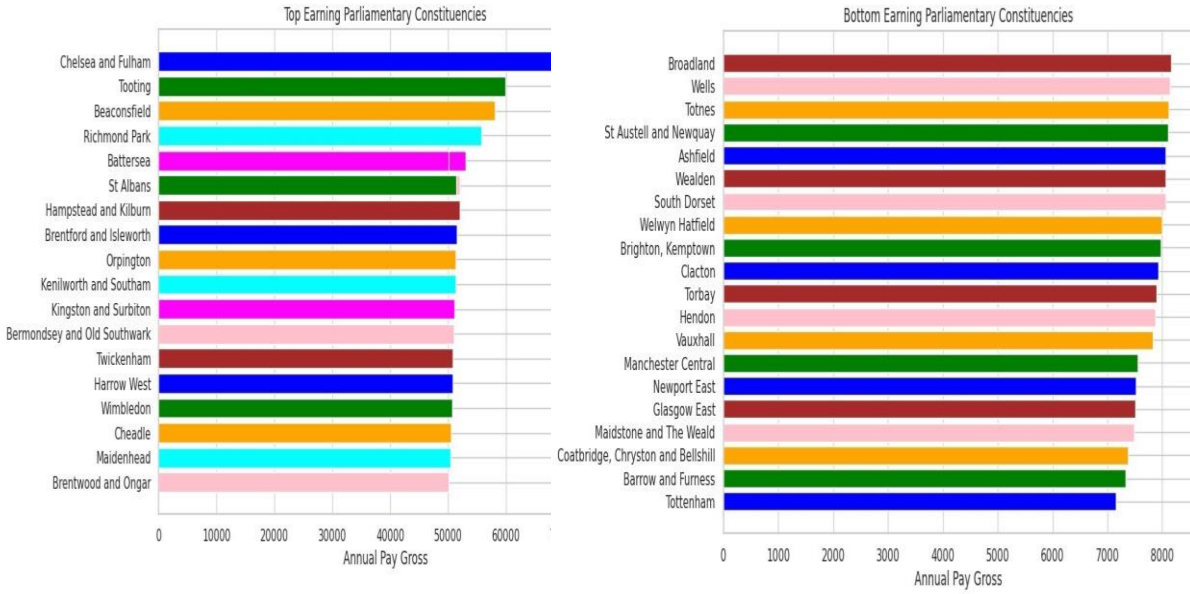

Figure 7: Wage variation by constituencies.

Model Performance Metrics

The table below showcases the calculated performance metrics for each model:

Table 1: Performance metric for each model.

Model	Mean Absolute Error (MAE) (£)	Root Mean Square Error (RMSE) (£)	R-squared (R2)
Linear Regression	3471.36	4629.43	0.592
Support Vector Machine	3760.70	5522.50	0.420
Random Forest	2942.13	4540.76	0.611
Ridge Regression	3517.28	4742.94	0.576
XGBoost	3475.43	4630.45	0.590

Model Performance Comparison

Visual representation of the models' performance (figure 8) across various metrics using bar charts:

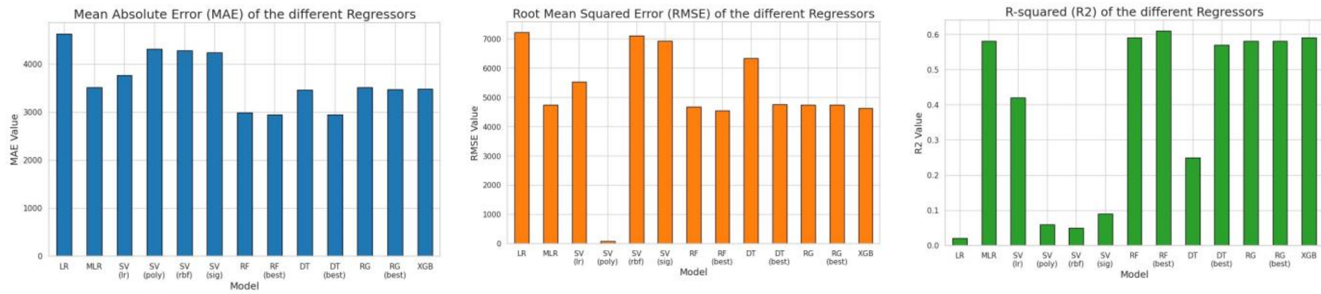

Figure 8: Model performance across metrics using bar chart.

Predictive Analysis

Random Forest: Demonstrates the strongest predictive performance, showcasing the lowest MAE and RMSE, signifying its ability to predict future wages with remarkable accuracy. The higher R-squared value reinforces its superior fit to the actual data.

XGBoost: Shows competitive performance metrics, closely following Random Forest, indicating robust predictive capabilities. Linear Regression: despite exhibiting reasonably accurate predictions, it registers slightly higher MAE and RMSE compared to Random Forest and XGBoost, suggesting a relatively less accurate prediction capability.

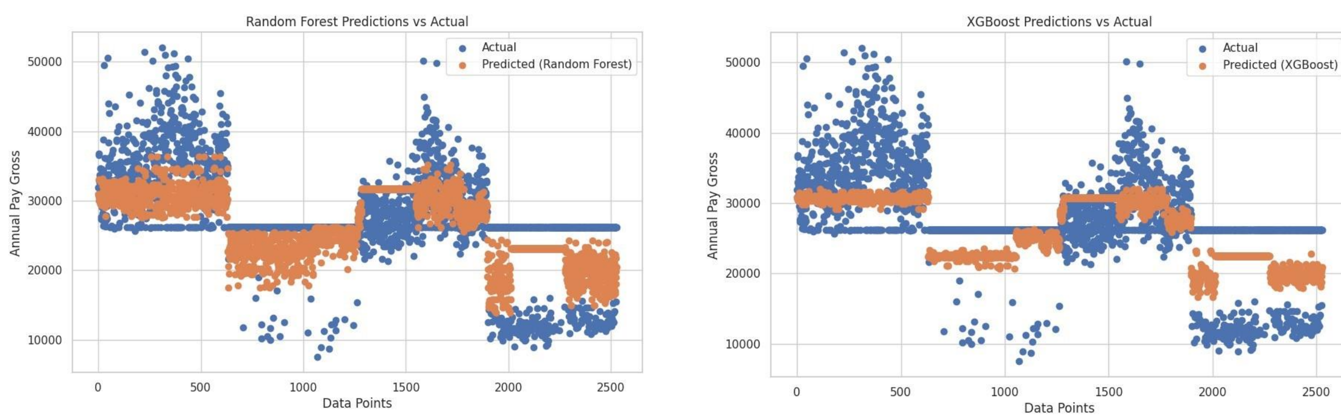

Figure 9: Scatter plots for predicted vs actual values.

The plot (figure 9) illustrates a pattern where both the random forest model’s predictions and randomly assigned values for annual pay gross seem to concentrate around £26,000, indicative of a prevalent cluster in the dataset. This implies that a significant portion of instances or predictions tends toward this income range, reflecting a distinct trend centred around middle-class earnings.

Predicted Annual Wages for Subsequent Years

The random forest model used several key factors, including hourly pay, total hours worked, gender, year, work type, educational qualifications, and age groups, as input features to predict annual wages. The predictions for annual wages over the next five years (2023-2027) are as follows:

Table 2: Model prediction over the next five years (2023 - 2027)

Year	Predicted Annual Wage
2023	28021.83
2024	28631.61
2025	29241.38
2026	29851.16
2027	30460.93

These prediction figures represent the anticipated annual wages based on the developed model's analysis and predictions. Based on the trend, it appears that people are expected to earn more in the future. The predicted annual wages show a consistent increase over the years from 2023 to 2027.

Discussion

The analysis delved into the intricacies of UK employee annual wage predictions by employing a diverse set of regression models. The Linear Regression model illuminated direct linear associations, offering a foundational understanding of wage determinants. Support Vector Regression (SVR), with its various kernels, revealed discrete non-linear patterns, showcasing the adaptability of the model to diverse data structures. The application of Random Forest Regression, utilizing ensemble learning, provided a comprehensive perspective by aggregating insights from multiple decision trees. Ridge Regression, by mitigating overfitting, ensured a balance between model complexity and performance. Random forest, using RandomizedSearchCV, systematically testing multiple hyperparameter combinations like the number of estimators, maximum depth, and minimum samples split demonstrated superior predictive power.

The significance of these models lies in their collective ability to capture the multiple aspects of the nature of wage fluctuations, reflecting the intricate dynamics within the UK labour market. The understanding derived from these models serves as a valuable resource for policymakers, enabling them to formulate targeted strategies for enhancing wage conditions in the workforce.

Comparing the models unveiled their distinct strengths and weaknesses. While Linear Regression presented a straightforward approach, it may oversimplify complex relationships within the data.

SVR, with its various kernels, showcased flexibility in handling diverse data structures, allowing for a more tailored analysis. Random Forest Regression excelled with its robust performance metrics, emerging as the most accurate model, yielding an RMSE of £4547.34, MAE of £2949.04, and an R2 of 0.6101. This signifies its exceptional predictive capabilities, outperforming other models in capturing wage dynamics.

Understanding these fine distinctions facilitates informed model selection, considering specific research objectives and dataset characteristics. The balance between model complexity and predictive performance should be carefully considered based on analytical requirements.

Despite their robustness, the models encountered certain unexpected limitations. Outliers, especially in higher wage brackets, challenged model predictions, indicating potential influential factors beyond the current feature set as seen from table 1 where RMSE values are higher than MAE values. Additionally, limitations in data details, particularly in certain demographic attributes, may impact prediction accuracy.

Our results align with existing literature, collaborating with broader discussions on wage dynamics, showcasing the relevance of employing a diverse suite of models to capture the intricate dynamics of the UK labour market. Through this analysis, factors like hourly pay, work type, and gender emerged as key influencers of annual wages. The models predict an upward trend in future earnings, suggesting a likely increase.

Conclusion

The investigation into UK employee annual wage dynamics, employing a diverse ensemble of regression models, signifies a comprehensive and methodological approach within the intersection of data science and economic analysis. This study has adeptly unveiled intricate relationships between predictor variables and annual wages, yielding invaluable insights vital for policymakers and economists in making effective labour market policies and societal frameworks.

The application of a diverse types of regression models has provided an in-depth understanding of the strengths and limitations inherent in predicting annual wages. While Linear Regression established basis for fundamental correlations, its inherent simplicity requires caution against oversimplifying the dynamic nature of wage determinants. Support Vector Regression (SVR) showed discrete non-linear patterns, demonstrating adaptability to diverse data structures. Ridge Regression carefully balanced model complexity, while XGBoost Regression showcased good predictive accuracy, particularly in capturing intricate wage dynamics. Random Forest Regression, through its ensemble learning methodology, presented the best and most comprehensive perspective, despite potential computational demands. Within the scope of these achievements, these models were faced with challenges. Handling outliers within higher wage brackets emerged as a notable challenge as seen in figure 5 and the tangible difference between RMSE values and MAE values as seen in table 1, signifying potential influential factors not wholly encompassed in the current feature set. Additionally, data limitations such as missing entries in columns (about 3,059) in specific demographic attributes posed challenge to achieving higher prediction accuracies. Complementing these findings, this research presents broader discussions on wage dynamics, demonstrating the significance of employing diverse models to comprehensively capture the intricacies of the UK labour market.

The interpretation of evaluation metrics, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and R-squared (R2), played a pivotal role in explaining model performances, offering insights into their predictive capabilities and limitations. The Random Forest model emerges as the most accurate, boasting an RMSE of 4547.34, an MAE of 2949.04, and an R2 of 0.6101, asserting its superiority in predictive accuracy.

Furthermore, the future wage predictions from 2023 to 2027, forecasted by the Random Forest model, provide crucial foresight into prospective annual wages:

- 2023: £28021.72
- 2024: £28631.49
- 2025: £29241.27
- 2026: £29851.04
- 2027: £30460.81

This study addresses the pivotal importance of delving into advanced predictive modelling techniques and cultivating a clear understanding of the dynamic nature of wage determinants. The implications extend beyond academic field, offering actionable insights such as wage regulation, devising incentive structures to boost different fields, optimal hiring strategies, skills development initiatives, enhanced business strategy, to policymakers, economists, and stakeholders. Acknowledging the existence of outliers and recognizing the ongoing necessity for refining models remains paramount. Given the challenges identified, future research endeavours should focus on refining models to handle outliers in higher wage brackets, incorporating additional relevant features like job experience, industry specific variables for enhanced predictive accuracy, examining temporal trends to capture evolving wage dynamics, and assessing model fairness across diverse demographic groups. These endeavours collectively aim for more accurate predictions and a thorough comprehension of wage dynamics in the UK labour market.

References

Abel, W., R. Burnham, & M. Corder (2016), "Wages, productivity and the changing composition of the UK workforce", Quarterly Bulletin 2016 Q1, pp. 12-22, Bank of England.

Alsulami, H., 2018. The effect of education and experience on wages: the case study of Saudi Arabia. American Journal of Industrial and Business Management, 8(1), pp.129-142.

Bank of England. (2016). Quarterly Bulletin pre-release article: Wages, productivity and the changing composition of the UK workforce. Retrieved from https://www.bankofengland.co.uk//media/boe/files/news/2016/march/qbpre-release-article-wages-productivity-and-the-changing-composition.pdf Accessed: 12th December 2023)

Bhattarai, K. and Wisniewski, T., 2017. Determinants of Wages and Labour Supply in the UK. Chinese Business Review, 16(3), pp.126-140.

Bowlus, A. J., & Grogan, L. (2009). Gender Wage Differentials, Job Search, and Part-Time Employment in the UK. Oxford Economic Papers, 61(2), 275– 303. Retrieved from JSTOR, http://www.jstor.org/stable/20529418 Accessed: 16th December 2023)

Buchanan, I., Pratt, A., & Francis-Devine, B. (2023). Women and the UK economy. Commons Library Research Briefing. Retrieved from https://researchbriefings.files.parliament.uk/documents/SN06838/SN06838.pdf Accessed: 16th December 2023)

Chen, L., Sun, Y., & Thakuriah, P. (2019). Modelling and Predicting Individual Salaries in United Kingdom with Graph Convolutional Network. In Hybrid Intelligent Systems: HIS 2018 (pp. 61–74). Advances in Intelligent Systems and Computing (Vol. 923). doi:10.1007/978-3-030-16042-5_6

Cribb, J., Joyce, R., & Wernham, T. (2022). Twenty-five years of income inequality in Britain: the role of wages, household earnings and redistribution. Institute for Fiscal Studies.

De Almeida Vilares, H., & Reis, H. (2022). Who's got the power? Wage determination and its resilience in the Great Recession. Centre for Economic Performance, London School of Economics and Political Science. (No. 1885). ISSN 2042-2695.

Dolton, P., Hantzsche, A., & Kara, A. (2020). The Dynamics of Public and Private Sector Wages, Pay Settlements and Employment. National Institute of Economic and Social Research.

Emerson, J. W., Green, W. A., Schoerke, B., & Crowley, J. (2013). "The Generalized Pairs Plot". Journal of Computational and Graphical Statistics, 22(1), 79–91. doi:10.1080/10618600.2012.694762.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). "Extremely Randomized Trees." Machine Learning, 63, 3-42. DOI: 10.1007/s10994-006-6226-1.

Kiss, Á., & Van Herck, K. (2019). Short-Term and Long-Term Determinants of Moderate Wage Growth in the EU. IZA Policy Paper No. 144. Institute of Labour Economics.

Koeniger, W., Leonardi, M., & Nunziata, L. (2007). Labor Market Institutions and Wage Inequality. Industrial and Labor Relations Review, 60(3), 340–356. Retrieved from http://www.jstor.org/stable/25249090 Accessed: 12th December 2023)

HM Treasury (2020) 'Impact of COVID-19 on working household incomes: distributional analysis as of May 2020'. [Online] Available at: https:// assets.publishing.service.gov.uk/media/5f0570673a6f40041295c1a5/ Impact_of_COVID-19_on_working_household_incomes.pdf (Accessed: 19th December 2023). ISBN 978-1-913635-39-8 PU 2975.

Mayhew, K. & Anand, P. (2020) 'COVID-19 and the UK labour market', Oxford Review of Economic Policy, 36, pp. S215-S224. doi: 10.1093/oxrep/graa017.

Saleh, A.K. Md. Ehsanes, Arashi, Mohammad, & Kibria, B.M. Golam. (2019). Theory of Ridge Regression Estimation with Applications. Hoboken: John Wiley & Sons, Inc.

Smola, A., & Schölkopf, B. (2004). A tutorial on support vector regression. S t a t i s t i c s a n d C o m p u t i n g , 1 4 , 1 9 9 - 2 2 2 . d o i : 1 0 . 1 0 2 3 /B%3ASTCO.0000035301.49549.88.

tastingbritain.co.uk. (2023). What was middle-class income in the UK in 2022? Retrieved from https://blog.moneyfarm.com/en/financial-planning/averaged i s p o s a b l e - i n c o m e - i n - t h e - u k - p e r - m o n t h / #:~:text=has%20been%20deducted.-,What%20was%20middle%2Dclass%20income%20in%20the%20UK%20in%202022,was%20the%20minimum%20for% 20London.

Wilkinson, L., & Friendly, M. (May 2009). "The History of the Cluster Heat Map". The American Statistician, 63(2), 179–184. doi:10.1198/tas.2009.0033.