Dean&Francis

ISSN 2959-6157

Construction and Application of Machine Learning Models for Salary Evaluation

Yutong Chen

School of Foreign Languages, Beihang University, Beijing, China
23379288@buaa.edu.cn

Abstract:

In the contemporary knowledge economy, data-driven salary evaluation is a strategic imperative for talent acquisition, retention, and equitable compensation. This paper presents a systematic review of the construction and application of machine learning models for salary prediction, charting the methodological evolution from the restrictive assumptions of traditional econometric paradigms to the superior predictive power of modern algorithms. To provide a holistic analysis, this study introduces a novel five-dimensional framework—"Data-Feature-Model-Explanation-Governance"—that moves beyond model accuracy to encompass the entire system lifecycle. The paper reviews state-of-the-art supervised learning models, including tree-based ensembles (e.g., XGBoost) and deep learning architectures (e.g., TabNet), and examines their diverse applications in empowering individuals, optimizing corporate HR strategies, and informing macroeconomic policy. A core finding is that the central challenge in the field has shifted from the pursuit of predictive accuracy to the imperative of building trustworthy AI systems. Consequently, the paper critically analyzes the indispensable roles of data governance, model interpretability (e.g., SHAP), and algorithmic fairness in mitigating bias and ensuring responsible deployment. It concludes by synthesizing these insights into a practical roadmap for researchers and practitioners, aiming to foster the development of transparent, equitable, and intelligent compensation systems.

Keywords: Salary Prediction; Machine Learning; Compensation Management; Human Resources Analytics; Explainable AI.

1. Introduction

1.1 Research Background

In the context of the knowledge economy, salary balance has become the core driving force for enhancing organizational competitiveness. The ability to retain outstanding talents depends on the fairness and transparency of the compensation system. Salary design is no longer a simple administrative issue, but a strategic focus of human resource management. Empirical based compensation decisions provide analytical basis for enterprises, help set competitive compensation ranges, optimize human capital investment, and provide reference for individuals to confirm their value in the market.

Salary prediction is a complex and profound task, and the formation of personal income depends on the interaction of multidimensional factors such as education level, work experience, skill level, geographical location, and industry attributes. This complexity has driven the development of methods, gradually expanding from traditional econometric models that rely on strict assumptions to more flexible and interpretable machine learning frameworks. The evolution of this model reflects a deeper understanding of compensation issues in both academia and industry and also indicates that human resource management is moving towards data-driven and intelligent direction.

1.1.1 The Limitations of Traditional Econometric Paradigms

Labor economics has dominated wage determination analysis for many years, and researchers often rely on statistical methods such as multiple linear regression (MLR). This model is highly valued for its ease of use and clear explanation of coefficients. Essentially, it uses ordinary least squares to fit linear equations to explain changes in the dependent variable. Its application is limited by data conditions, and linear assumptions often do not match the actual salary distribution. The change pattern of experience return rate is obviously nonlinear. OLS estimation also shows a high sensitivity to outliers, and a small group of high paying individuals may weaken the overall prediction reliability, making it difficult for the model to meet practical needs. To address specific issues such as sampling selection bias, researchers have developed the Heckman correction model. This method has value in causal inference, but it is based on strict normality assumptions and requires appropriate instrumental variables as support, which is often difficult to achieve in practical research [1,2]. The evolution of the model reflects the limitations of traditional methods in terms of explanatory power and applicability and also highlights the complexity inherent in salary prediction problems.

1.1.2 The Paradigm Shift to Machine Learning

The transition to machine learning (ML) in salary prediction is underpinned by demonstrated empirical superiority and fundamental theoretical advantages. ML models consistently exhibit superior predictive performance, a differential that stems from a fundamental divergence in modeling objectives [3]. Econometric models were primarily designed for explanatory inference, prioritizing the impact of a limited set of pre-specified variables. Machine learning models, in contrast, are optimized for predictive accuracy on unseen data. They are architected to autonomously discover complex, non-linear patterns and high-order interaction effects within high-dimensional feature spaces. For example, the value of a particular skill is highly contextual, interacting with experience, industry, and location. While linear models fail to capture this context without extensive manual engineering, tree-based ensembles and neural networks are inherently designed to model such compositional effects [4,5,6]. This capacity for automated pattern recognition renders ML models fundamentally better aligned with the complex realities of wage determination.

1.1.3 Diversified Application Scenarios

The application of these advanced predictive models is reshaping compensation management across multiple societal strata. For individuals, public-facing AI-powered salary calculators are democratizing access to compensation data and mitigating the information asymmetry that has historically favored employers. For corporations, ML models are being integrated into core human resources workflows to drive intelligent compensation strategies, including generating equitable salary recommendations, conducting systemic pay equity audits to rectify wage gaps and proactively identifying attrition risks [7]. At the macroeconomic level, these models serve as powerful new instruments for analyzing labor market dynamics and simulating the potential impacts of policy interventions with unprecedented granularity [8].

1.2 Research Objectives and Significance

Given the rapid technological advancements and the expanding scope of applications, a systematic examination of the construction and application of machine learning models for salary evaluation is both timely and necessary. This paper seeks to move beyond a simple survey of algorithms to provide a structured and holistic analysis of the end-to-end process.

1.2.1 Core Objectives

This study undertakes a systematic review of the algorith-

ISSN 2959-6157

mic evolution, data ecosystems, and practical deployment of salary prediction models. It aims to chart the methodological trajectory from foundational econometrics to state-of-the-art machine learning, critically analyze the key components of the modeling pipeline, and investigate the challenges and solutions related to responsible deployment, with a specific focus on interpretability and fairness. These findings are then synthesized into a coherent guiding framework for academic inquiry and industry practice.

1.2.2 A Five-Dimensional Analytical Framework

To achieve these objectives, this paper introduces and applies a five-dimensional analytical framework titled "Data-Feature-Model-Explanation-Governance." This framework posits that a robust and responsible system must be assessed holistically, not on model accuracy alone. The framework integrates five core dimensions: Data, the foundational layer concerning sources and quality; Features, the transformation of raw data into meaningful predictors; Model, the selection and validation of the predictive algorithm; Explanation, techniques that render model decisions transparent; and Governance, the operational and ethical oversight addressing challenges like bias and compliance. This structure provides a comprehensive lens through which to analyze the field and propose a roadmap for future development.

1.2.3 Contribution and Structure

This paper aims to provide a definitive theoretical reference and a practical implementation roadmap for researchers and practitioners. By systematically organizing the vast body of knowledge through the proposed five-dimensional framework, this work endeavors to bridge the gap between algorithmic theory and applied HR strategy. The remainder of this essay is structured as follows. Chapter 2 reviews the theoretical foundations of both traditional and machine learning models. Chapter 3 explores the application of these models across personal, corporate, and macroeconomic domains. Chapter 4 delves into critical challenges and future directions for the field. Finally, Chapter 5 presents a concluding synthesis of the key findings.

2. Theoretical Foundations of Salary Prediction and Machine Learning

The analytical methodologies for salary evaluation have evolved significantly, shifting from assumption-laden econometric models designed for inference to a diverse ecosystem of machine learning algorithms optimized for predictive accuracy [9]. This chapter outlines this theoretical landscape, contrasting foundational statistical methods with the machine learning paradigms that now define the state of the art.

2.1 Traditional Salary Prediction Methods and Their Limitations

For decades, wage analysis was the province of labor economics, which relied on statistical tools whose applicability is limited by restrictive assumptions that are often violated by modern, high-dimensional HR data.

2.1.1 Foundational Regression Models

Multiple Linear Regression (MLR) has long been the benchmark, modeling salary as a linear function of various predictors. The model's appeal lies in the straightforward interpretability of its coefficients, which are typically estimated via Ordinary Least Squares (OLS). However, the validity of MLR is contingent upon stringent assumptions, most notably a linear relationship between predictors and the outcome. This is frequently not the case, as the effect of experience on salary is demonstrably curvilinear. While robust alternatives like Quantile Regression can model different points of the salary distribution and are less sensitive to outliers, they generally still assume linearity.

2.1.2 The Heckman Correction Model

The Heckman correction model was specifically developed to address sample selection bias, which arises because wage data is only observed for employed individuals, who are not a random sample of the population [1,10]. This two-stage model first estimates the probability of being in the sample and then uses a derived correction term to adjust the main wage equation for this bias. While theoretically elegant, its practical application is constrained by strong statistical assumptions, including joint normality of error terms, and the often-difficult requirement of finding a valid instrumental variable. These traditional models, while useful for structural inference, are ill-suited for the modern goal of maximizing predictive accuracy in complex feature spaces.

2.2 Machine Learning Models

The shift to machine learning marks a transition from an explanatory to a predictive modeling culture. Free from the restrictive parametric assumptions of their econometric predecessors, ML algorithms excel at learning complex patterns directly from data.

2.2.1 Supervised Learning

In supervised learning, the core of salary prediction, an algorithm learns a mapping function from input features to a known salary outcome. Ensemble methods using decision trees have become the industry standard for such structured, tabular data. A Random Forest (RF)constructs numerous decorrelated decision trees and averages their predictions, a process that mitigates overfitting and enhances generalization [11]. In contrast to RF's parallel approach, Gradient Boosted Decision Trees (GBDT) build trees sequentially, with each new tree trained to correct the errors of the prior ensemble. Prominent, highly optimized GBDT implementations like XGBoost and LightGBM often deliver state-of-the-art accuracy and are renowned for their performance and efficiency [12].

While tree ensembles dominate, deep learning architectures are an emerging frontier. The Multi-Layer Perceptron (MLP), a foundational neural network, can approximate any continuous function, enabling it to model highly complex, non-linear relationships, though it requires large datasets and can be difficult to interpret [13]. More specialized architectures like TabNet have been designed specifically for tabular data, incorporating a sequential attention mechanism that offers a degree of interpretability alongside high performance. Furthermore, Transformer-based models are being adapted to learn rich, contextual embeddings for categorical features, promising further advances.

2.2.2 Unsupervised and Representation Learning

Unsupervised methods are also employed to discover latent structures within HR data, providing valuable context for compensation strategy. Clustering algorithms such as K-Means and Gaussian Mixture Models are used to segment jobs into distinct levels or create data-driven employee personas for compensation benchmarking. An emerging and more advanced approach involves Graph Neural Networks (GCN), which move beyond treating employees as independent instances. By modeling the network of relationships between employees, a GCN can learn representations that encode not only an individual's attributes but also their critical relational context within the organization.

Table 1. A Comparative Analysis of Supervised Learning Models for Salary Prediction

Model	Core Mechanism	Primary Strengths	Common Limitations
Random Forest	Ensemble of decorrelated decision trees via bagging and feature subsampling.	High accuracy, robustness to overfitting, native handling of mixed data types.	Can be computationally intensive; individual trees are not directly interpretable.
GBDT (XGBoost/LightGBM)	Sequential, additive ensemble of trees trained to correct prior errors.	Often state-of-the-art accuracy; regularization controls overfitting; highly efficient implementations.	More sensitive to hyperparameter tuning than RF; can overfit on noisy data.
MLP	Feedforward neural network with hidden layers and non-linear activations.	Universal function approximator; capable of learning highly complex patterns; automated feature engineering.	Requires large data, computationally expensive, "black box" nature, prone to overfitting.
TabNet	Deep learning with a sequential attention mechanism for feature selection.	Combines high performance with a degree of built-in interpretability.	More recent and less established than GBDTs; performance can be dataset-dependent.
GCN	Neural network operating on graph structures, aggregating neighborhood information.	Explicitly models relational data, capturing structural context beyond instance features.	Requires data to be structured as a graph; methodology is more complex.
Stacking	A meta-model learns to combine predictions from a diverse set of base models.	Often achieves higher accuracy than any single constituent model.	Complex to implement correctly; computationally expensive to train.

2.2.3 Hybrid and Ensemble Strategies

To maximize predictive performance, practitioners often employ hybrid strategies that combine the outputs of multiple, diverse models, leveraging the principle that different algorithms make different types of errors. Stacking, or stacked generalization, is a sophisticated ensemble technique where a meta-model is trained on the out-of-fold predictions of several base models, such as an XGBoost model and an MLP. This allows the meta-model to learn the optimal weighting of each base predictor, often yielding an accuracy greater than any single constituent model. To prevent information leakage, predictions on the training data are generated via a k-fold cross-validation procedure. Blending is a simpler and computationally less expensive variant of stacking, where the training data is split into a training set and a small validation set, and the

ISSN 2959-6157

base models are trained on the former, with their predictions on the latter used to train the meta-model. While faster, blending can be more prone to overfitting if the validation set is not representative.Table1 provides a comparative analysis of various supervised learning models used for salary prediction. It outlines the core mechanisms, primary strengths, and common limitations of each model, including Random Forest, GBDT (XGBoost/LightGBM), MLP, TabNet, GCN, and Stacking. This table serves as a valuable reference for understanding the capabilities and trade-offs of different algorithms in the context of predictive analytics.

3 Domains of Application and Real-World Case Studies

The theoretical framework and techniques outlined in the last chapter are not simply the result of intellectual discourse but are being actively applied to a variety of practical uses, generally reshaping compensation administration and understanding. This application is occurring at three distinct levels: providing career management techniques to citizens, establishing corporate human resources departments as strategic, analytical divisions, and empowering policymakers with fresh tools for examining the labor market. This chapter bridges the theory-to-practice gap by exploring some of the most significant applications and case studies within each of these areas.

3.1 Personal Salary Tools

The most direct manifestation of machine learning in terms of compensation is the development of consumer oriented online websites aimed at enhancing the capabilities of labor market employees. These websites utilize predictive models to provide personalized salary insights, thereby eliminating information asymmetry that historically favored employers in salary negotiations.

3.1.1 Online Salary Calculators and Negotiation Aids

The AI salary calculator serves as a market intelligence tool for job seekers, relying on data from employers, voluntary submissions from users, and online job postings to train the model. They can present a complete salary range, help individuals compare their income with market levels, and also measure the economic returns of investing in new skills. In this way, salary information has become more widespread, and historical data inequality has been weakened. In recent years, these tools have gradually evolved into negotiation aids, combining career achievements with industry data to provide factual evidence for salary adjustments. Its credibility is limited, especially when data relies on self-reported information with inflation bias, which is more suitable as a link in negotiations rather than the sole basis. This change indicates that data-driven tools can enhance individuals‘ bargaining confidence, but in practical applications, caution should still be exercised regarding the source and quality of data.

3.2 Corporate HR Management

Within organizations, machine learning is being integrated directly into human resources information systems (HRIS) and workflows to render compensation decisions more systematic, strategic, and equitable. This integration signifies a fundamental shift from a reactive and often adhoc approach to salary administration towards a proactive, data-driven framework of „intelligent compensation."

3.2.1 New Hire Salary Recommendations and Benchmarking

Corporations increasingly deploy internal models to generate consistent, equitable salary recommendations for new hires. This practice serves the dual objectives of attracting elite talent with competitive offers while maintaining internal equity. The core goal of such implementations is to leverage data to facilitate the „most equitable and least biased salary decisions possible," thereby mitigating the influence of cognitive biases that can pervade manual compensation setting.

3.2.2 Identifying and Mitigating Systemic Pay Gaps

Arguably the most critical corporate application is in the execution of pay equity audits. AI-powered platforms conduct sophisticated statistical analyses of internal HR data to identify compensation disparities between demographic groups that are not explained by legitimate, job-related factors. This enables organizations to proactively rectify systemic inequities, as exemplified by Salesforce‘s widely-cited initiative to address its gender pay gap [14]. These systems are crucial for preventing models from learning and perpetuating historical biases present in the training data.

3.2.3 Linking Compensation to Employee Attrition

Salary is a recognized driving factor for retaining employees, and machine learning models are now widely used to predict which employees are at the highest risk of voluntary turnover. By identifying high-performance employees whose salaries are lower than the market median or internal peers, these models can trigger positive intervention measures. This insight enables targeted salary adjustments to retain key talent, transforming compensation from lagging indicators to strategic retention levers. However, this type of enterprise adoption creates a complex dynamic that may lead to ‚monitoring compensation‘, where opaque algorithmic systems constrain employees to less transparent forms of management.

3.3 Macroeconomic and Policy Analysis

The application of machine learning in salary analysis has expanded beyond individual and corporate levels to include macroeconomic research and public policy evaluation. With the help of models, researchers can predict the overall trend of the labor market, simulate the effects of policy interventions in more detail, and provide new tools for macro level judgments.

3.3.1 Forecasting Salary and Labor Market Trends

Time series prediction models, especially long short-term memory networks (LSTM), are suitable for capturing dynamic changes in the labor market. Based on large-scale continuous datasets, such as online job information, such systems can accurately predict salary trends, skill demands, and employment levels. For policy makers and economic planners, these forecasts not only provide forward-looking indicators, but also reveal potential directions for structural changes. More importantly, this predictive mechanism reflects the practical value of data-driven policy tools in addressing complex market environments.

3.3.2 Policy Simulation

The role of machine learning in policy simulation is becoming increasingly prominent, especially in the study of the impact of minimum wage increases on employment. By relying on more flexible modeling methods, a processing group with a wider coverage can be constructed to more accurately reflect the situation of the affected population. A study covering 172 minimum wage increases in the United States shows that affected workers‘ wages have significantly increased, while employment levels have not shown a significant decline. This result challenges the conclusions of some traditional economic models and demonstrates the comprehensive explanatory power of machine learning beyond causal inference. For public decision-making, such methods suggest that policy makers should not rely solely on classical models, but should consider intelligent tools to make policy evaluations closer to the complex face of the actual labor market.

4 Critical Challenges and Future Directions

The predictive function of machine learning in salary analysis has been fully validated, but its effective and ethical application depends on solving a series of core problems beyond the scope of model performance. The implementation of artificial intelligence in sensitive and critical areas such as employee compensation requires a high level of attention to reliable and responsible practices. The long-term effectiveness and trust of stakeholders not only depend on the reduction of prediction errors, but also on stable data governance, fair algorithm design, and system construction with interpretability and transparency. These issues touch upon the fundamental aspects of institutions and values, rather than just technical issues. This chapter focuses on these challenges and points out future research directions, with the goal of establishing a more just and intelligent compensation system that truly serves the common development of organizations and individuals through technological progress.

4.1 Data Governance and Privacy

The operation of machine learning systems depends on the integrity and quality of the underlying data. In the field of human resources, this type of data is highly sensitive and requires governance models centered around privacy protection, security, and lifecycle management. If the governance structure is insufficient, it will not only weaken the effectiveness of the model, but may also trigger legal risks and trust crises. For enterprises, data governance is not simply a technical configuration, but a strategic choice that determines whether compensation analysis can maintain fairness and reliability in the long run.

4.1.1 Data Privacy and Security

The human resources dataset contains a large amount of personal identity information, and collecting this information for model training may pose serious privacy and security risks. The industry is gradually adopting privacy preserving machine learning (PPML) solutions to reduce risks. Federated Learning (FL) has established a decentralized architecture where models are trained locally on distributed data sources and only perform centralized anonymous parameter updates to avoid storing raw employee data in a centralized manner [15]. Differential privacy (DP) relies on mathematical mechanisms to add statistical noise during the computation process while maintaining personal privacy and ensuring overall statistical utility. These technologies indicate that governance issues are not just technical details, but the key to establishing trust mechanisms.

4.1.2 Data Quality and Lifecycle Management

The accuracy of the salary model depends on the quality of data and the adaptability of the lifecycle in dynamic environments. Human resource data often has missing information, and using simple imputation methods can distort the distribution. Multiple imputation (MI) utilizes a more rigorous process to generate multiple complete datasets to reflect the uncertainty of missing values. The labor

ISSN 2959-6157

market environment is non-stationary, and the interaction between predictive factors and compensation will drift over time, which requires the model to have dynamism. Continuous monitoring and periodic retraining have become necessary conditions to keep the model consistent with the real economy. This process demonstrates that data management is not only a supporting tool, but also the cornerstone for ensuring the long-term effectiveness of predictive systems.

4.2 Model Interpretability and Fairness

In high-risk compensation decisions, the comprehensibility and fairness of models have become the most prominent ethical issues. This leads to two directions: explainable artificial intelligence (XAI) and algorithmic fairness.

4.2.1 Explainable AI (XAI)

High performance models such as gradient boosting trees and neural networks often exhibit „black box" characteristics, lack comprehensibility, and are prone to damaging the trust and compliance of stakeholders. Explainable artificial intelligence methods aim to provide explanations for the inference process of models. SHAP (Shapley Addition Interpretation) relies on game theory to reveal the impact of individual predictions and demonstrate the overall operational logic of the model at the global level [3]. Transparency is not only beneficial for debugging and auditing but also helps to explain to employees the basis for the formation of salary results, thereby enhancing acceptance.

4.2.2 Algorithmic Fairness

The most significant risk in salary prediction is that the model learns and amplifies social biases in historical data, thereby solidifying the salary gap. Researchers have proposed a series of methods to alleviate this problem. The preprocessing stage can correct the training data and reduce bias; During the modeling phase, fairness constraints can be introduced into the objective function; The post-processing stage can adjust the model output to meet fairness standards. These processes reveal that privacy, interpretability, and fairness are not isolated issues, but intertwined pillars of responsible artificial intelligence.

4.3 The Next Frontier

The field of salary analysis is rapidly developing, and new research directions are shaping the future salary management system. This includes shifting from correlation research to causal research to reveal the true impact of factors such as professional certification on compensation pathways; Multimodal data fusion combines unstructured information with structured human resource data to generate more comprehensive predictive models; Reinforcement learning driven dynamic compensation planning enables compensation to no longer rely on annual evaluations, but to adapt in real-time to changes in individuals and the market. These explorations suggest that the compensation system may evolve towards personalization, intelligence, and continuous adjustment. Table 2 is a summary of a framework for responsible AI in compensation analytics.

Table 2. A Framework for Responsible AI in Compensation Analytics

Challenge Area	Specific Problem	Proposed Solution/Technique	Core Principle
Data Governance & Privacy	Centralization of sensitive HR data creates security and privacy risks.	Federated Learning (FL): Train models decentrally without moving raw data.	Data Minimization & Security
Data Governance & Privacy	Incomplete datasets lead to biased or inefficient models.	Multiple Imputation (MI): Generate multiple plausible datasets to account for uncertainty.	Statistical Robustness
Data Governance & Privacy	Model performance degrades over time due to changing market conditions.	Concept Drift Detection & Retraining: Monitor data distributions and periodically retrain the model.	Adaptability & Lifecycle Management
Model Interpretability	Complex models like GBDT and DNN hide their decision-making logic.	SHAP (SHapley Additive exPlanations): Use game theory to fairly attribute prediction impact to each feature.	Transparency & Accountability
Algorithmic Fairness	Model learns and perpetuates historical biases from training data.	Bias Mitigation Techniques (Pre-, In-, Post-processing): Modify data, algorithm, or predictions to satisfy fairness constraints.	Equity & Non-discrimination

5 Conclusion

The application of machine learning to the evaluation and prediction of salary represents a profound and accelerating transformation in the way individuals, corporations, and policymakers understand and manage compensation. This technology has catalyzed a decisive departure from the constraints of traditional statistical models, ushering in a new paradigm of data-driven prediction that offers superior accuracy and a capacity to model the true, multifaceted complexity of the modern labor market. This paper has systematically charted the evolution of this field, elucidated its theoretical underpinnings, surveyed its practical applications, and critically examined the challenges that must be surmounted for its responsible deployment.

5.1 Research Summary

This review has traced the clear analytical evolution from interpretable yet rigid econometric frameworks, such as Multiple Linear Regression and the Heckman correction model, to the high-performance, non-parametric tree-based ensembles like Random Forest and XGBoost that define the current industry standard.The field is now advancing into a new frontier of deep learning architectures, including TabNet and Transformers, alongside relational methods like Graph Convolutional Networks that offer deeper structural insights. To structure this analysis, this paper introduced and applied a five-dimensional „Data-Feature-Model-Explanation-Governance" framework, ensuring a holistic assessment that moves beyond a narrow focus on model performance to encompass the entire system lifecycle.

5.2 Core Insights

A primary insight from this review is the established dominance of tree-based ensembles, particularly gradient boosting implementations, for structured compensation data due to their ability to capture complex, non-linear interactions. However, the most critical insight is that the central challenge confronting the field is no longer solely the pursuit of predictive accuracy but the imperative of building trustworthy AI systems. The future of intelligent compensation depends on the holistic development of systems that are demonstrably fair, transparent, and accountable. This requires an integrated approach that embeds privacy-preserving technologies like federated learning, systematically utilizes XAI techniques such as SHAP for transparency, and rigorously audits for and mitigates algorithmic bias.

5.3 Practical Significance and Future Outlook

The findings of this paper offer methodological guidance for key stakeholders, providing governments with more accurate tools for labor market analysis, equipping corporations to build strategic and equitable compensation frameworks, and empowering individuals with tools for career navigation. Looking forward, progress will be defined by advancements on several research frontiers. There is a pressing need for standardized benchmark datasets to facilitate reproducible research. The next wave of innovation will likely emerge from the intersection of domains, such as developing federated causal inference techniques. Ultimately, the field must evolve from static models towards real-time, incremental learning systems that can adapt dynamically to both individual performance and macroeconomic shifts. In conclusion, while the methodologies to build more efficient and equitable compensation systems now exist, the defining challenge—and greatest opportunity—lies not in further algorithmic invention, but in their thoughtful and responsible implementation to foster a more transparent and just labor market for all.

References

[1] Heckman J J. Sample selection bias as a specification error. Econometrica: Journal of the econometric society, 1979: 153-161.

[2] Yang S. Automated employee salary prediction algorithm based on machine learning//International Conference on Computer Vision, Application, and Algorithm (CVAA 2022). SPIE, 2023, 12613: 243-249.

[3] Lothe D M, Tiwari P, Patil N, et al. Salary prediction using machine learning. INTERNATIONAL JOURNAL, 2021, 6(5).

[4] Matbouli Y T, Alghamdi S M. Statistical machine learning regression models for salary prediction featuring economy wide activities and occupations. Information, 2022, 13(10): 495.

[5] Arik S Ö, Pfister T. Tabnet: Attentive interpretable tabular learning//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(8): 6679-6687.

[6] Kablaoui R, Salman A. Machine learning models for salary prediction dataset using python//2022 International Conference on Electrical and Computing Technologies and Applications (ICECTA). IEEE, 2022: 143-147.

[7] Zhu J. Unveiling Salary Trends: Exploring Machine Learning Models for Predicting Data Science Job Salaries//2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024). Atlantis Press, 2024: 173-182.

[8] Dustmann C, Lindner A, Schönberg U, et al. Reallocation effects of the minimum wage. The Quarterly Journal of Economics, 2022, 137(1): 267-328.

[9] Wolpert D H. Stacked generalization. Neural networks, 1992,

ISSN 2959-6157

5(2): 241-259.
[10] Song X, Mitnitski A, Cox J, et al. Comparison of machine learning techniques with classical statistical models in predicting health outcomes//MEDINFO 2004. IOS Press, 2004: 736-740.
[11] Asaduzzaman A, Uddin M R, Woldeyes Y, et al. A Novel Salary Prediction System Using Machine Learning Techniques//2024 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON). IEEE, 2024: 38-43.
[12] Quan T Z, Raheem M. Salary prediction in data science field using specialized skills and job benefits–a literature. Journal of Applied Technology and Innovation, 2022, 6(3): 70-74.
[13] Prasetio A B, bin Mohd Aboobaider B, bin Ahmad A. Machine Learning for Wage Growth Prediction: Analyzing the Role of Experience, Education, and Union Membership in Workforce Earnings Using Gradient Boosting. Artificial Intelligence in Learning, 2025, 1(2): 153-173.
[14] Madhani P M. Salesforce Control and Compensation System: A Game Theory Model Approach. Compensation & Benefits Review, 2015, 47(4): 190-202.
[15] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data// Artificial intelligence and statistics. PMLR, 2017: 1273-1282.