# MULTI-MODAL ENSEMBLE LEARNING ARCHITECTURE FOR LABOR MARKET'S DIGITAL TRANSFORMATION

**Assylbek Gizatov**
Senior Research Scientist, Research Institute of Mathematics and Applied Technologies, Atyrau University named after Kh. Dosmukhamedov, Kazakhstan
e-mail: a.gizatov@asu.edu.kz | ORCID: 0009-0009-2557-0767

**Nurtas Zhaksybayev**
Master's student in Applied Data Analytics, School of AI and Data Science, Astana IT University, Kazakhstan
e-mail: 255235@astanait.edu.kz | ORCID: 0009-0008-1031-3795

**Alibi Jangeldin**
M.Sc in Statistics, Senior-Lecturer, Department of Computing and Data Science, Astana IT University, Kazakhstan
ORCID: 0009-0000-0990-4362

**Abstract**

The rapid growth of educational grants in Kazakhstan has led to wage stagnation and oversaturation in key sectors. This study presents a meta-model combining CatBoost, Explainable Boosting Machine (EBM), and Bayesian-optimized LSTM neural networks to forecast wages, analyze labor market factors, and predict sector demand trends. Findings highlight critical oversupply risks, providing policymakers actionable insights for aligning education with economic needs, fostering sustainable development.

## Keywords

Kazakhstan, labor market, wage stagnation, educational grants, machine learning, CatBoost, Explainable Boosting Machine (EBM), Long Short-Term Memory (LSTM), Bayesian optimization, demand forecasting, economic sustainability.

## 1. Introduction

The rapid increase in educational grants in Kazakhstan has intensified competition in its labor market, causing wage stagnation and oversaturation in key sectors [14, 18]. Effective forecasting is essential to align educational outputs with market demands and support sustainable economic growth [1, 18]. Advanced ensemble methods, particularly CatBoost, efficiently manage categorical data [4, 13], while data transformations such as Box-Cox and

Yeo-Johnson improve predictive accuracy [2, 17]. Additionally, Microsoft's Explainable Boosting Machine (EBM) provides interpretability by clarifying key predictive features [3, 11]. For temporal forecasting, Long Short-Term Memory (LSTM) networks, optimized through Bayesian methods, effectively capture sequential data dependencies [5, 15, 16]. This study integrates these methodologies into a comprehensive meta-model, offering actionable insights to policymakers and educational institutions.

## 2. Methods and Methodology

This chapter outlines the methodological framework employed in the study, designed to forecast labor market trends and analyze risks of market oversaturation. The proposed methodology integrates data processing, modeling techniques, and performance evaluation methods within a structured meta-model framework. Each subsection details specific steps ranging from initial data description to the evaluation of model accuracy and performance.

### Data Description

The study utilizes a comprehensive dataset obtained from Kazakhstan's Bureau of National Statistics, specifically the "Report on the Structure and Distribution of Wages" (Form 2-T). This report, mandatory for organizations with over 250 employees, covers detailed employment metrics from 2019 to 2023. The dataset comprises over 2.6 million records with 11 key attributes: **Unique Employee ID, Gender and birth year, Hire date, Occupational classification, Education level, Payment type (tariff or non-tariff), Wage fund per employee, Work schedule (full-time or part-time), Hours worked.**

These variables collectively support comprehensive labor market analysis, enabling evaluation of wages, employment trends, and demographic shifts across sectors over multiple years. This robust dataset allows for both cross-sectional and longitudinal analyses crucial for accurate forecasting and identification of labor market trends (see *Figure 1*).

### Data Anonymization via Gaussian Copula-based Synthesis

To adhere to data privacy requirements while preserving statistical properties essential for analysis, data anonymization via Gaussian Copula-based synthesis was employed. This method involves modeling the joint distribution of features based on Sklar's Theorem:

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)), \qquad (1)$$

The Gaussian copula effectively captures the dependence structure among variables, enabling generation of synthetic datasets. The Gaussian copula is defined in Formula:

$$C_{\Sigma}^{\text{Gauss}}(u_1, u_2, \ldots, u_d) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_d)) \qquad (2)$$

with its density given by Formula:

$$C_{\Sigma}^{\text{Gauss}}(u_1, \ldots, u_d) = \frac{1}{|\Sigma|^{1/2}} \exp(-\frac{1}{2} z^T(\Sigma^{-1} - I)z) \qquad (3)$$

Synthetic data generated using this approach maintains the marginal distributions and inter-variable relationships of the original dataset, thus ensuring analytical validity while protecting individual privacy.

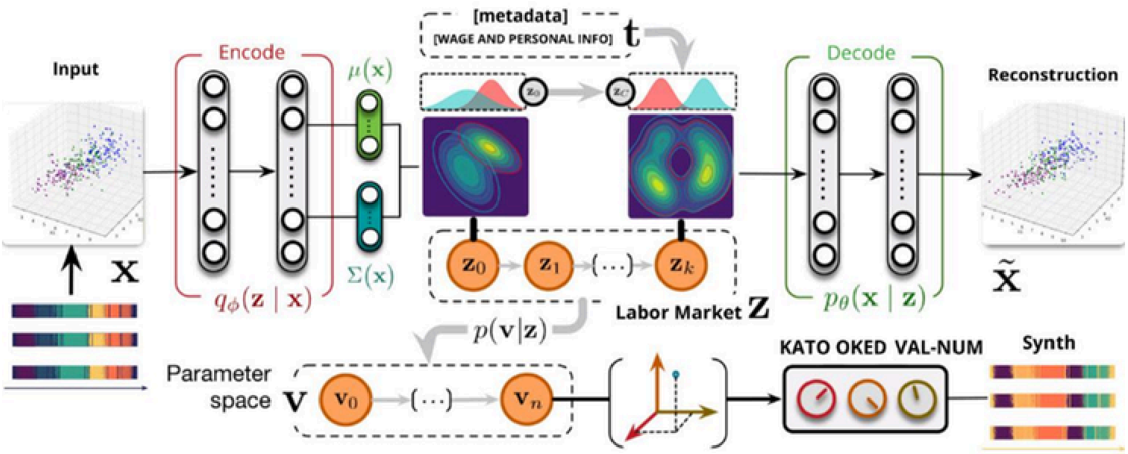


*Fig.1: Dataflow through Multivariate Analysis*

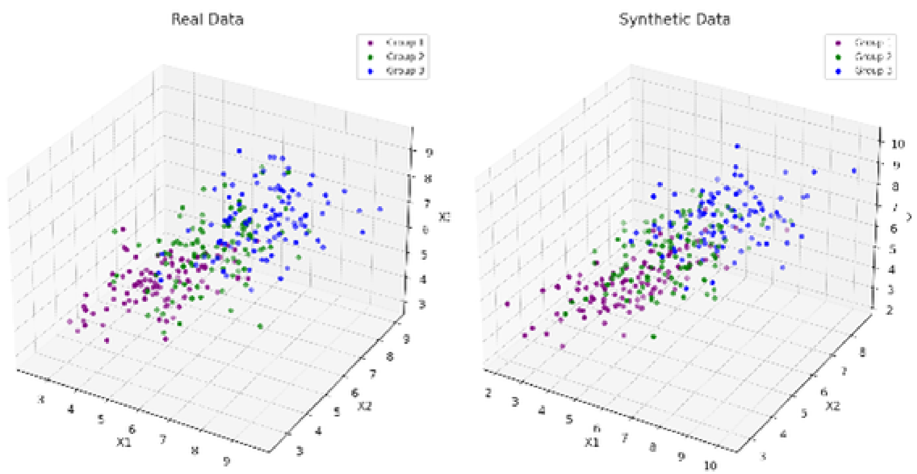


*Fig.2: Comparison of Real and Synthetic Data's correlations*

## Modeling Techniques:

### Yandex CatBoost model with Box-Cox and Yeo-Johnson methods

We selected the Yandex CatBoost model as the first model of the final ensemble meta-model, as our dataset has a large number of categorical features. Before the usage of the Yandex CatBoost model, numerical features within the input data matrix often undergo power transformations. These transformations are intended to control variance, make the distribution of data more similar to Gaussian, and increase the linearity of the data, leading to better (i.e., more stable) modeling results. The Box-Cox and Yeo-Johnson transformations, for example, are popular choices for these transformations.

$$y' = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, & if\ \lambda\neq0, \\ \log(\lambda), & if\ \lambda=0. \end{cases}, \qquad y' = \begin{cases} \frac{(y+c)^{\lambda}-1}{\lambda}, & if\ \lambda\neq0, \\ \log(y+c), & y\geq-c, \lambda=0, \\ \frac{(-y+c)^{(2-\lambda)}-1}{2-\lambda}, & y<-c, \lambda\neq2, \\ -\log(-y+c), & y<-c, \lambda=2. \end{cases} \tag{4}$$

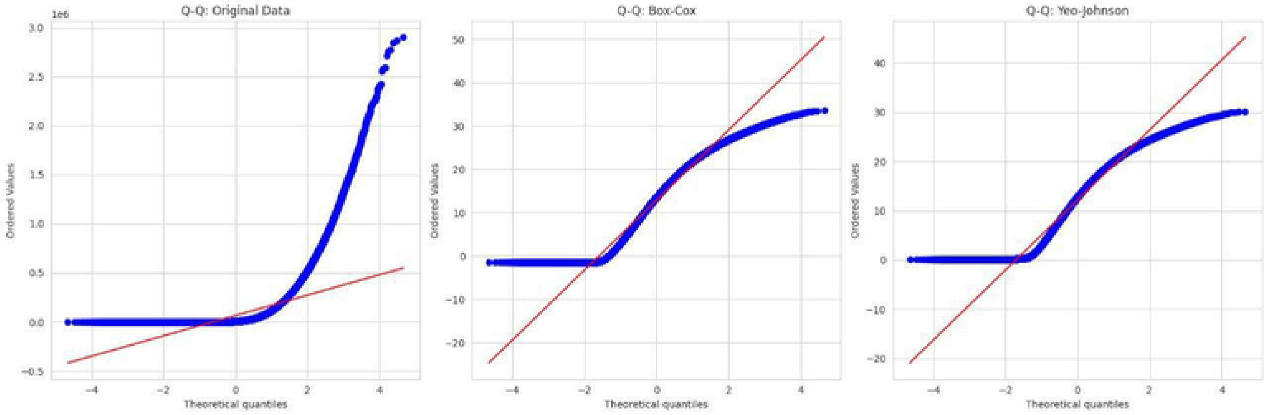


*Fig. 3: Variance stabilization*

The Yandex CatBoost algorithm was then applied to this feature matrix and now contains appropriately transformed numerical features (denoted X′). CatBoost is a decision tree-based gradient boosting algorithm, particularly optimized to handle categorical features, which it processes internally and efficiently.

$$f(x)=\sum_{m=1}^{M} \gamma_m T_m(x), \tag{5}$$

The model's decision function, represented as an ensemble of decision trees, is given by:

$$f(x)=T_1(x)+T_2(x)+\cdots+T_M(x), \tag{6}$$

where:

- $T_m(x)$ represents the prediction of the m-th decision tree.
- M is the total number of trees in the ensemble.

The iterative learning process of CatBoost follows the boosting approach:

$$T_m(x)=T_{m-1}(x)+\eta\cdot h_m(x), \tag{7}$$

where:

- η is the learning rate, controlling the contribution of each new tree.
- $h_m(x)$ is the new decision tree trained on the residuals.

This formulation demonstrates how CatBoost refines its predictions iteratively, combining multiple weak learners (decision trees) into a strong predictive model. CatBoost architecture and data flow demonstrate how input data are processed through the model.

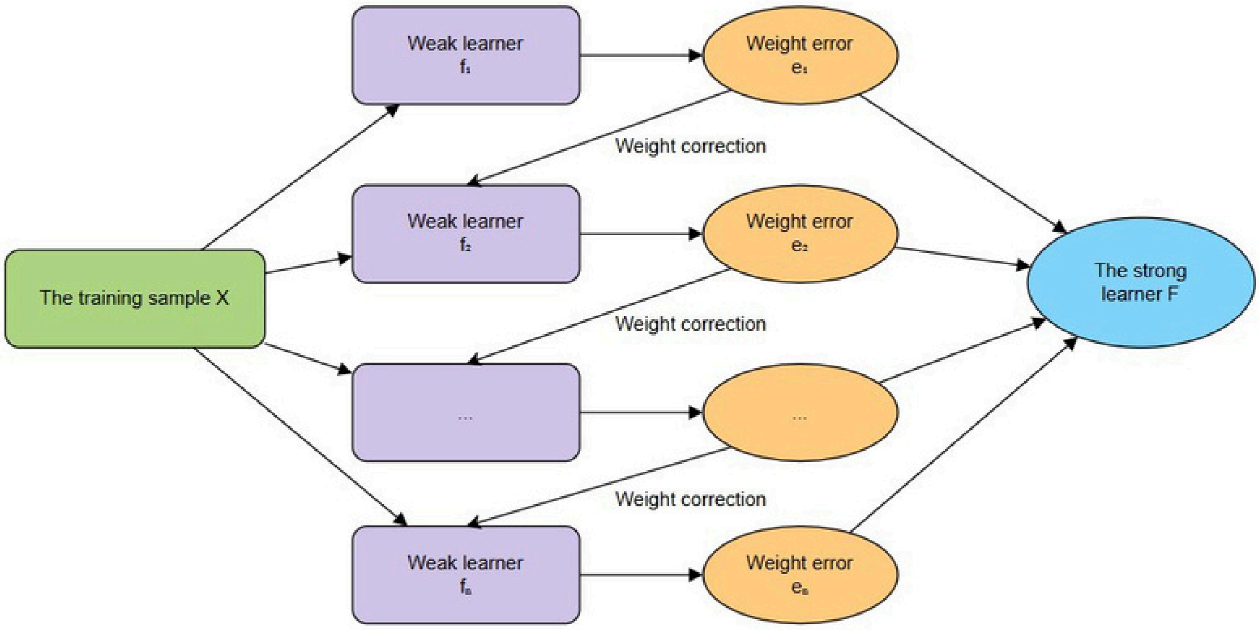


*Fig. 4: Architecture of Catboost model and its data flow*

This formulation demonstrates how CatBoost refines its predictions interatively, combining multiple weak learners (decision trees) into a strong predictive model. CatBoost architecture and data flow demonstrate how input data are processed through the model.

### Explainable Boosting Machine (EBM)

The Explainable Boosting Machine (EBM) model, proposed by Microsoft Research, was employed to estimate the importance of parameters and factors affecting the income of these professions. The Explainable Boosting Machine provides transparency and interpretability of model predictions, structured according to the **generalized additive model (GAM)** framework:

$$g(E[y])=\beta_0+\sum_{j} f_j(x_j)+\sum_{i\neq j} f_{ij}(x_i,x_j) \tag{8}$$

which sums the effects of individual features and select pairwise interactions. For each feature $xj$, the model learns a shape function $f_j(x_j)$, which represents the contribution of that feature to the output. Interaction terms $f_{ij}(x_i,x_j)$ are optionally included when they significantly improve model performance without compromising interpretability. The learning process in EBM is stage-wise and cyclic: it trains shallow trees for one feature at a time using gradient boosting, iteratively refining the feature's shape function. After the main effects are learned, EBM optionally incorporates interaction terms by identifying high-contributing feature pairs and fitting two-dimensional shape functions. This approach allows users and policymakers to trace prediction outcomes directly back to specific features, enabling transparent decision-making and accountability. Importantly, because EBM is additive and low-complexity, it avoids the opaqueness common to black-box models while still achieving competitive performance.

### Panel LSTM reinforced with Bayesian Optimization

We utilized a panel data representation of time dependence to model and predict time dependent demand dynamics across occupations using Long Short-Term Memory (LSTM) networks. The panel data method makes it possible to model multiple time series (i.e., demand for different professions across time) at the same time, so the model can incorporate both overall shared patterns across professions and individual patterns specific to professions.

$$f_t=\sigma(W_{(f\,[ht-1,\,xt])}+b_{(f)}), \tag{9}$$

$$i_t=\sigma(W_i\,[h_{t-1},x_t]+b_i) \tag{10}$$

$$\widetilde{c_t}=\tanh(W_C\,[h_{t-1},x_t]+b_c) \tag{11}$$

$$o_t=\sigma(W_O\,[h_{t-1},x_t]+b_o) \tag{12}$$

$$c_t=f_t\odot c_{t-1}+i_t\odot\widetilde{c_t} \tag{13}$$

$$h_t=o_t\odot\tanh(c_t) \tag{14}$$

In this study, a panel LSTM architecture was implemented to simultaneously forecast trends across multiple occupational time series. Each sequence, corresponding to a specific profession, was encoded using an embedding layer that captures profession-specific features. This allows the model to generalize patterns across professions while preserving individual differences. To improve predictive accuracy and training efficiency, Bayesian optimization:

$$x^*=\arg\min_{x\in\mathcal{X}} f(x) \tag{15}$$

was applied to fine-tune the model's hyperparameters. These included the embedding size, LSTM layer size, number of dense layers, dropout rates, learning rate, and batch size. The optimization process used a Gaussian Process-based surrogate model to explore the hyperparameter space efficiently. The training process was carefully monitored for convergence using early stopping criteria and validation loss tracking. *Figures 5 and 6* visualize the LSTM network's structure and the Bayesian optimization workflow, respectively. This combination of deep learning and statistical optimization allows the model to accurately capture and forecast dynamic labor market behaviors with minimal overfitting.

### The Meta-Model: A Synergistic Ensemble Framework

The proposed meta-model integrates the outputs of CatBoost, EBM, and LSTM models into a unified decision framework to leverage their respective strengths accuracy, interpretability, and temporal generalization. The architecture of the meta-model is illustrated in *Figure 7*.

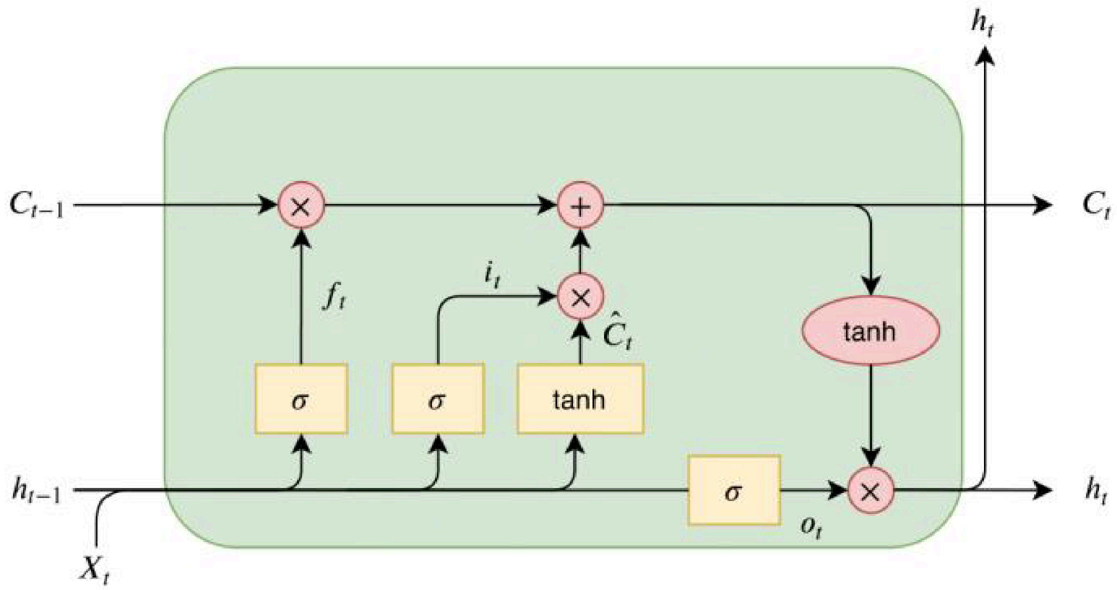


*Fig. 5: LSTM architecture*

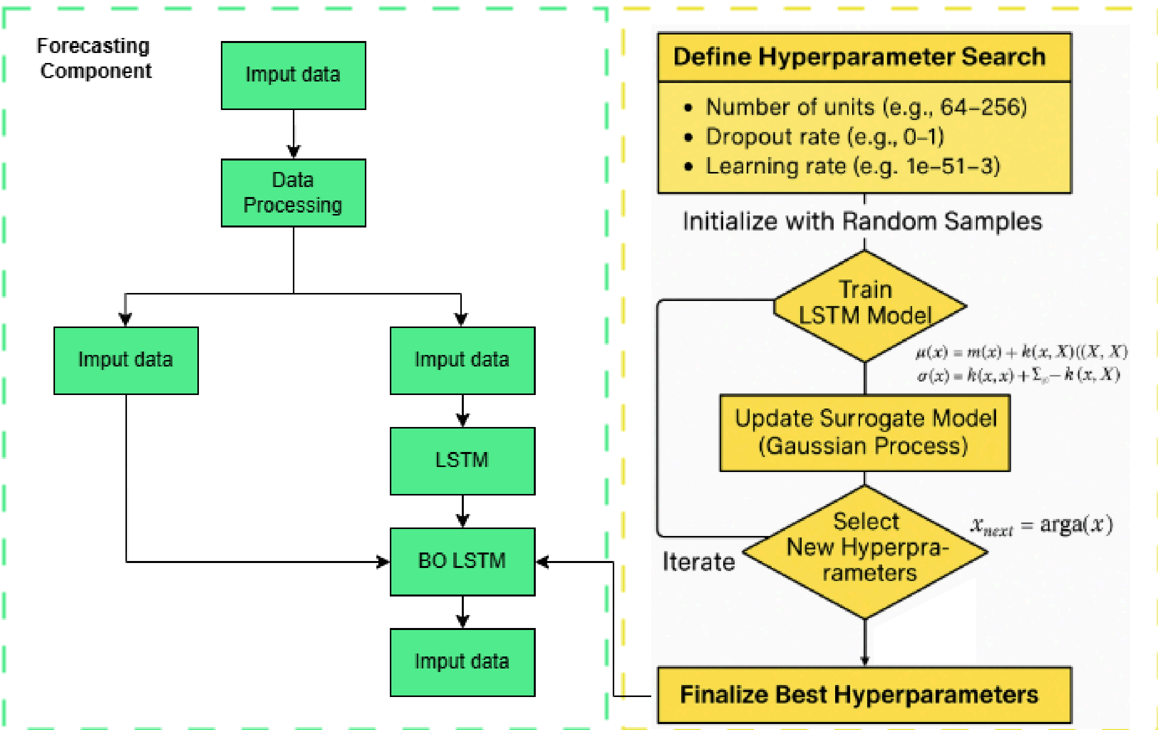


*Fig. 6: Bayesian optimization LSTM flow chart*

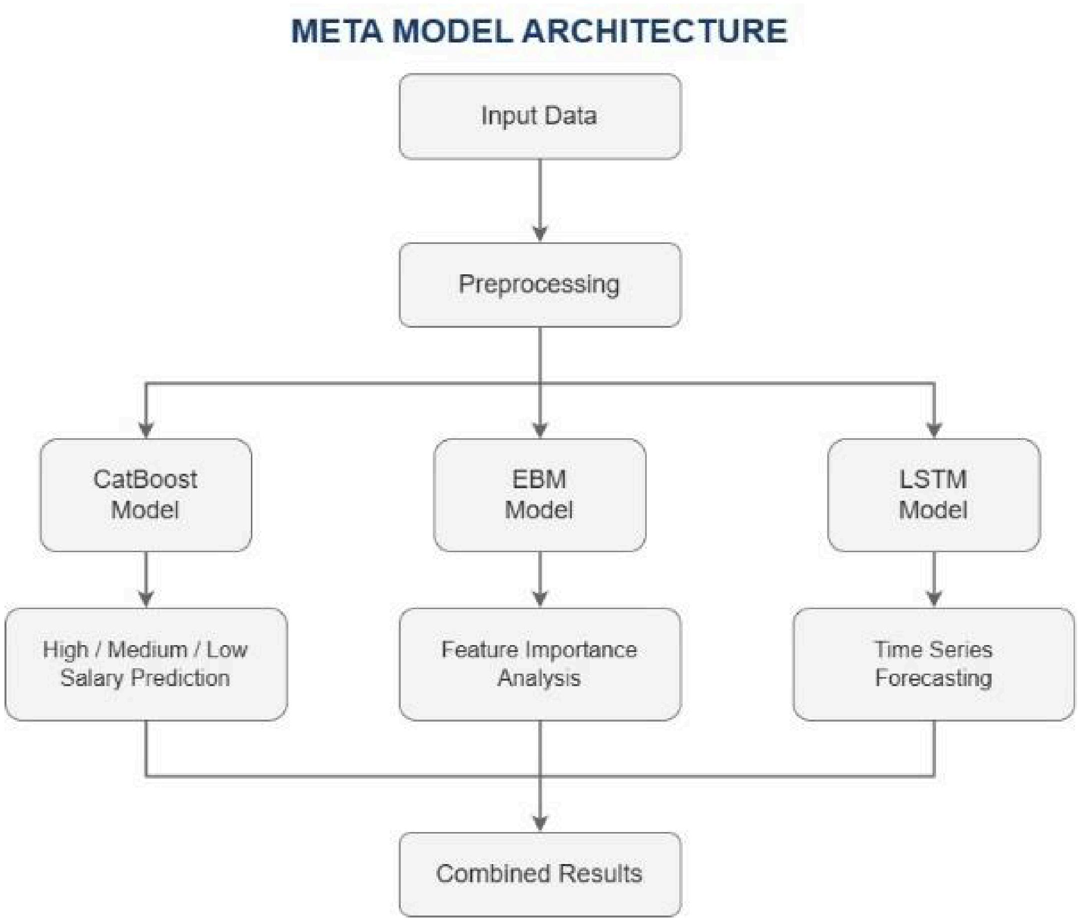


*Fig. 7: Architectural design of the Meta-Model framework.*

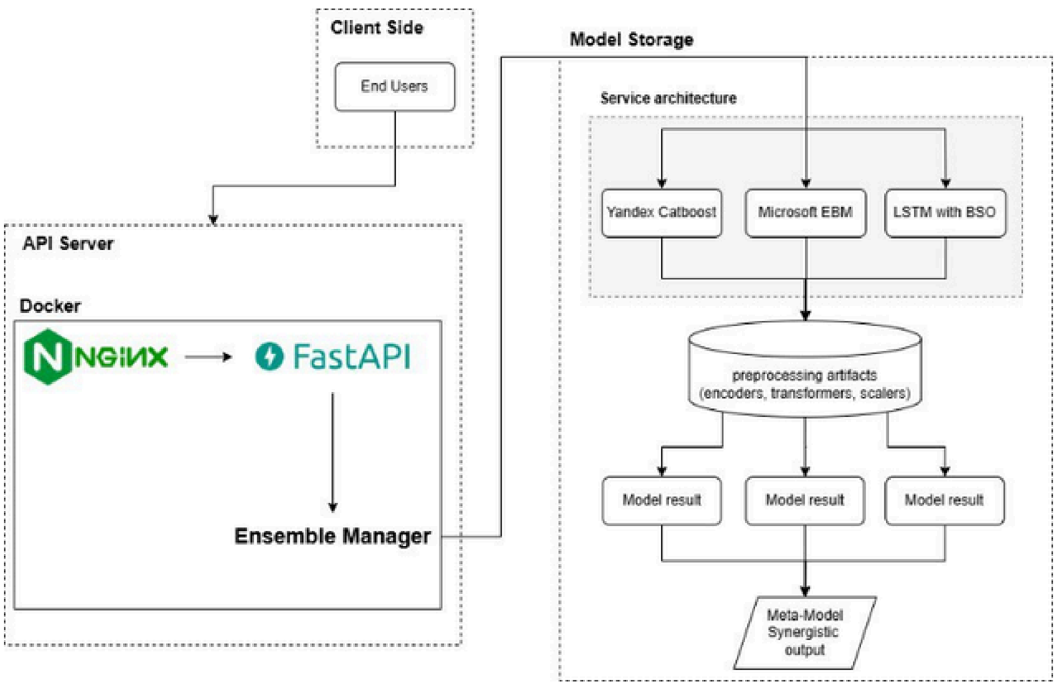


*Fig. 8: Deployment flow*

In this structure, input data is first preprocessed and then distributed across three parallel branches:

- The CatBoost model performs classification of salary levels (High/Medium/Low);
- The EBM model assesses feature importance and provides interpretable predictions;
- The LSTM model forecasts profession-specific labor demand over time.

Each model contributes its output to the final ensemble, which merges predictions using a weighted strategy optimized through cross-validation. This allows **the final synergistic output to reflect both statistical accuracy and interpretability.**

$$F(f_1(x_i),f_2(x_i),f_3(x_i),\alpha) = \alpha_1 f_1(x_i)+\alpha_2 f_2(x_i)+\alpha_3 f_3(x_i) \tag{16}$$

**Problem Formulation and Performance Evaluation.**

The primary goal of this study is to identify sector-specific trends in wage dynamics and employment saturation by constructing a predictive meta-model based on historical labor market data. The problem is formulated as a multi-output forecasting task, where classification (salary band prediction), regression (wage value prediction), and time series forecasting (labor demand trends) are performed simultaneously. To evaluate the performance of the CatBoost, EBM, and LSTM models, we applied three core metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) (Formula 2.24: Ошибки). These metrics provide a quantitative basis for comparing model accuracy across both regression and classification tasks.

$$MAE=\frac{1}{N}\sum_{1=1}^{N}(y_i-\hat{y}_i), \tag{17}$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}, \tag{18}$$

$$\mathrm{MAPE} = \left(\frac{100\ \%}{\mathrm{N}}\right)\sum_{i=1}^{N}\left|\frac{(y_i-\hat{y}_i)}{y_i}\right| \tag{19}$$

MAE captures the average magnitude of errors in the predictions without considering direction. MAPE expresses error as a percentage, making it easier to interpret in relative terms. RMSE gives higher weight to larger errors, offering a more sensitive measure for applications where outliers or larger deviations are critical. These three metrics together ensure a balanced assessment of model reliability, generalization, and performance under varying conditions. Through these formulations and evaluations, the proposed meta-model can be rigorously assessed and reliably used for policymaking and strategic workforce planning.

**CatBoost Modeling**

The CatBoost model shows strong predictive accuracy, especially for medium- and high-income groups, with predictions closely aligned to actual values. This indicates its effectiveness where sufficient data exists. Low-income predictions show more variance due to

## 3. Results and Discussions

fewer records and noise from synthetic data, which was partly mitigated by Box-Cox transformation. Despite this, CatBoost reliably uncovers hidden relationships between demographics, education, occupations, and wages, demonstrating robustness in modeling an imbalanced labor market.

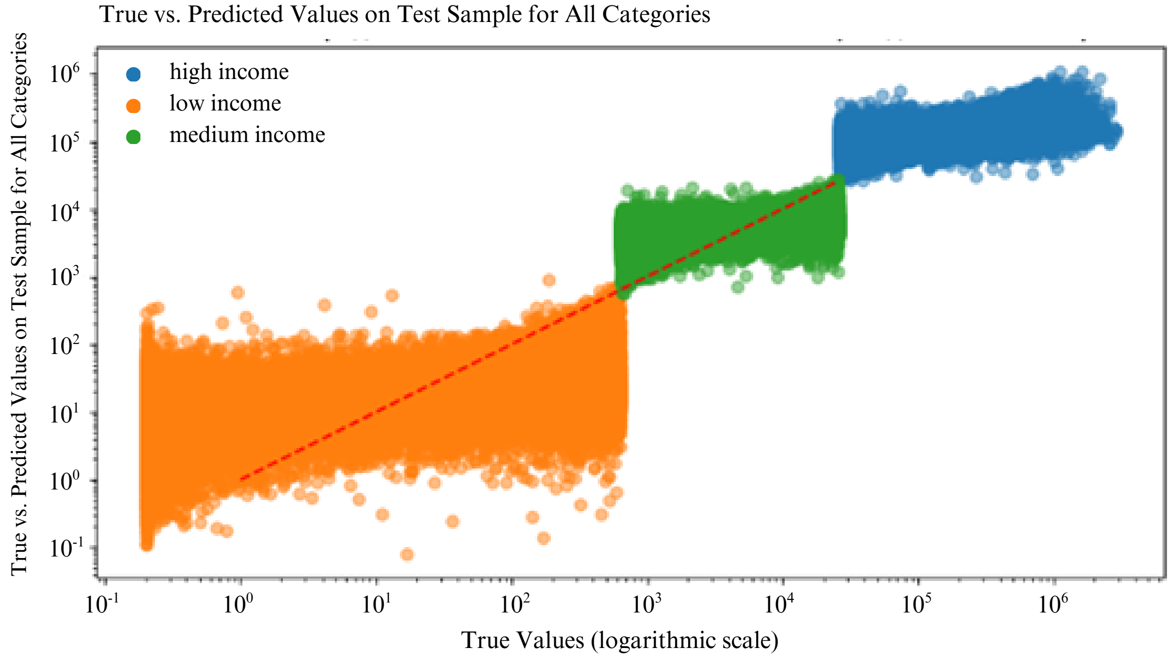


*Fig. 9: Catboost Perfomance*

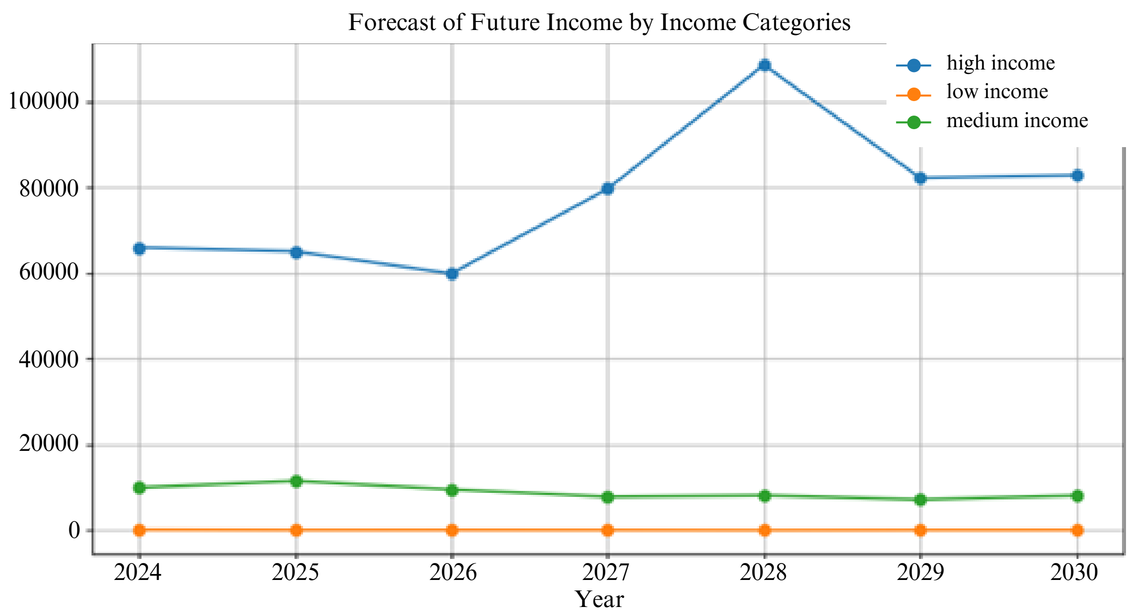


*Fig. 10: Income Forecasts through 2030*

Income forecasts through 2030 reveal key economic trends. High-income levels are expected to rise until 2026, then decline, indicating saturation in oversupplied fields like law, finance, and IT. Meanwhile, medium- and low-income groups show stagnant or slightly rising wages, risking declining real incomes amid inflation. These patterns highlight a structural imbalance: some sectors face wage compression and saturation, while others lack growth. This underscores the need for targeted labor policies in education and workforce planning. CatBoost's insights can guide adjustments in training, support for emerging industries, and correcting labor market imbalances.

### EBM Model: Feature Importance and Interpretability

The Explainable Boosting Machine (EBM) provides a transparent analysis of wage drivers through a combination of individual feature contributions and their interactions. It does this using a generalized additive model that is both statistically robust and human-interpretable. In our experiments, the model identified education level (SUO), region-specific GDP (GSDPRS), and industry classification (SPPN2) as the most critical predictors of income, as shown in the global importance chart (*Figure 11*).

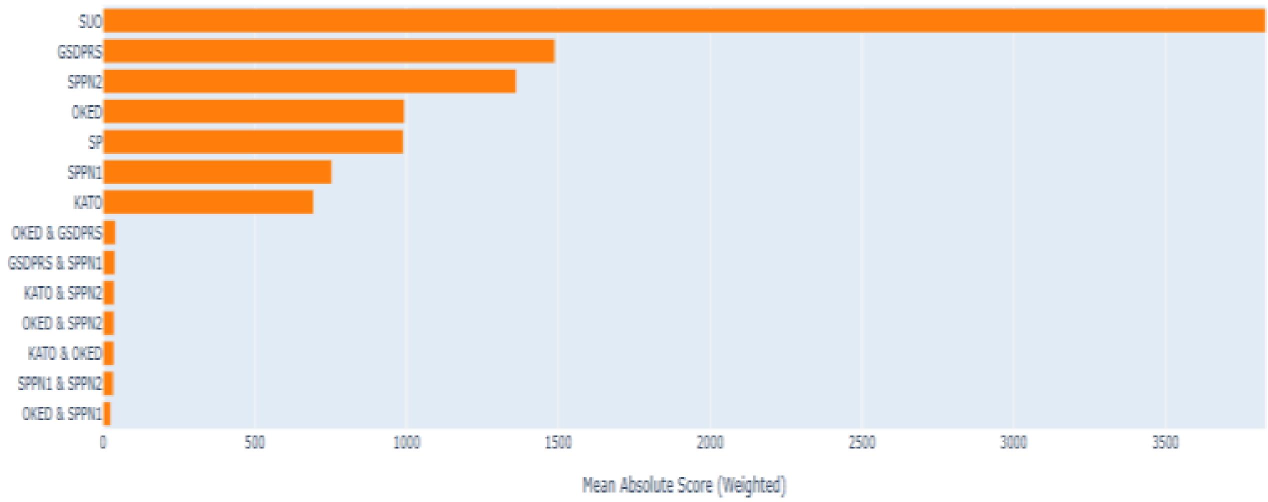


*Fig. 11: Global feature importance*

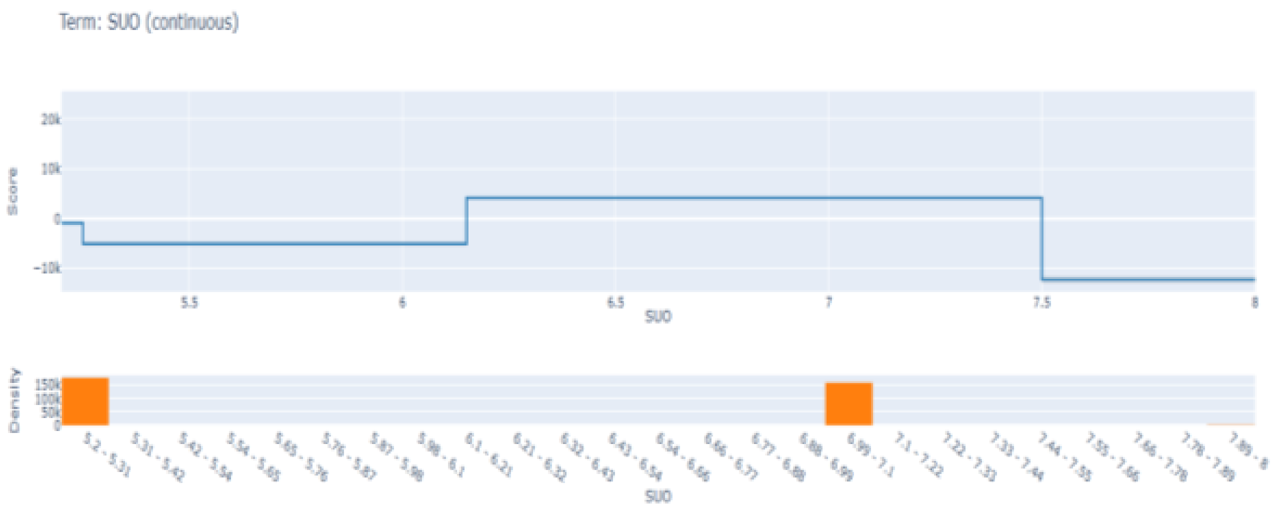


*Fig. 12: Partial-dependence plot for SUO*

*Figure 12* highlights the relationship between education (SUO) and income, revealing that wages increase with education up to a certain point, after which they plateau and in some cases decrease. This non-linear pattern suggests overqualification or inefficient matching between advanced education and job availability. It supports the case for more targeted funding strategies that consider optimal educational thresholds for specific sectors. *Figure 13* provides a local explanation for a single instance, showing how each input factor pushed the predicted wage up or down. The visual breakdown—such as positive contributions from GDP and industry type, and negative influences from certain regional codes—enables decision-makers to interpret forecasts in real-world terms. This helps improve trust in AI-driven policy tools, especially in sensitive domains like public sector employment and grant funding.

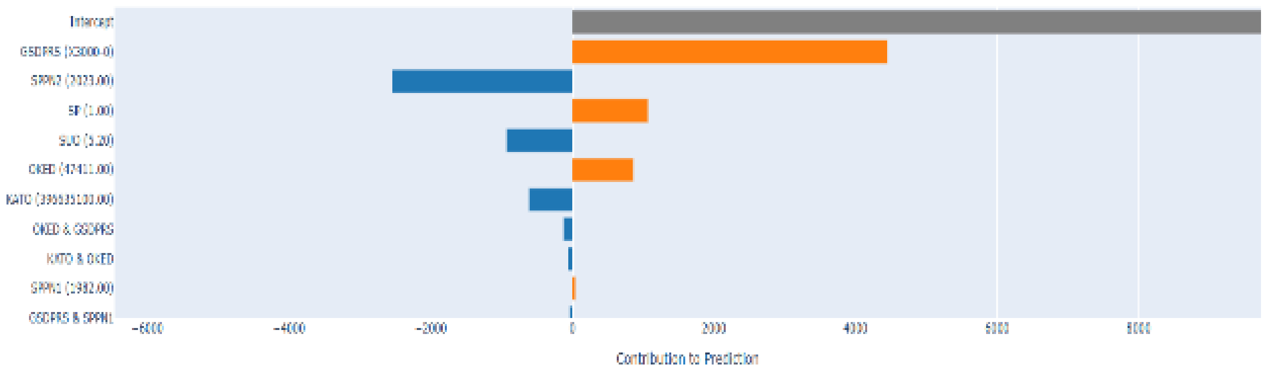


***Fig. 13: Partial-dependence plot for SUO***

**Panel Long-Short Term Memory Neural Network Reinforced by Bayesian Stochastic Optimization for Demand Forecasting**

For temporal forecasting of labor demand, the panel LSTM model provided accurate and stable predictions across multiple professions. By modeling demand sequences simultaneously and using shared representations, the panel LSTM captures profession-specific trends while leveraging inter-profession relationships. Hyperparameters were optimized through Bayesian methods, enabling convergence to high-performance configurations efficiently.

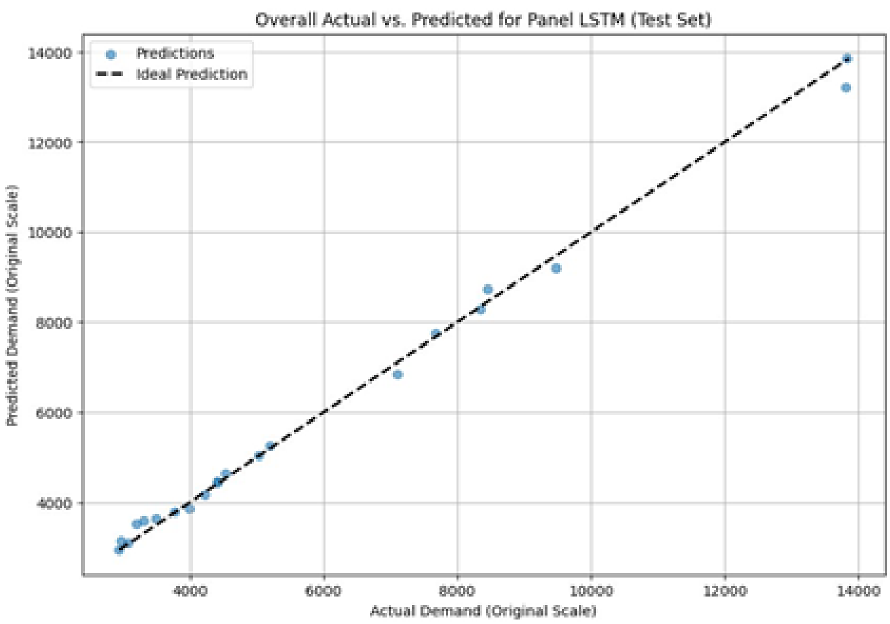


***Fig. 14: Overall actual vs. predicted demand***

The model's broader accuracy is confirmed in *Figure 14*, which plots actual versus predicted demand across all professions in the test set. The tight clustering along the ideal diagonal line indicates low prediction error and high correlation. Such consistency across professions affirms the model's scalability and relevance for national workforce planning. *Figure 15* depicts the demand forecast for profession 2350-3-003, illustrating strong agreement between the actual 2023 values and the model's prediction. The forecast suggests a tapering trend through 2026, potentially signaling labor saturation. These early warnings are essential for policymakers to adjust educational intake or reallocate grants to underrepresented sectors.

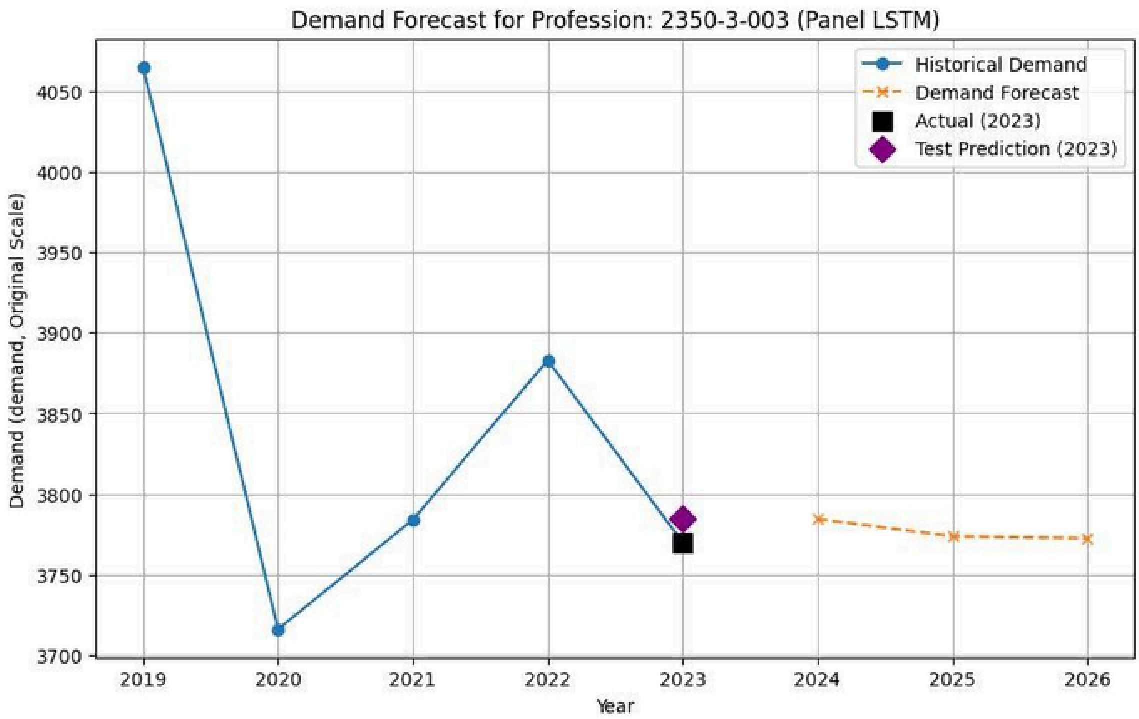


*Fig. 15: Historical demand and future forecast for a profession*

### Meta-Model: Synergistic Integration of Predictive Approaches

The final ensemble meta-model synergizes CatBoost's wage classification, EBM's interpretability, and LSTM's temporal forecasting into a coherent and modular architecture. This integration ensures the system can address wage analysis from multiple perspectives—static, interpretable, and time-evolving.

A practical demonstration of the ensemble's inference process is shown in *Figure 16*. For three example professions, the meta-model outputs wage forecasts, identifies contributing features (e.g., GDP, occupation codes), and highlights potential mismatches in demand. This fusion of quantitative and explanatory power is especially useful for ministries and educational institutions tasked with strategic workforce planning. In conclusion, the experiments validate the effectiveness of each model individually and highlight the advantages of their integration. The approach not only ensures high prediction accuracy but also provides explainable and actionable insights that are crucial for shaping effective labor and education policy.

```
Prediction Results:

CatBoost predictions (salaries in different sectors):
  - high: [733844.118003017, 457727.98066390754, 361598.41225407267]
  - medium: [250675.94031274103, 258546.3445583005, 263002.5711379661]
  - low: [99434.38870504327, 115405.74699270257, 62300.91483344173]

EBM predictions (factor importance for professions):
  - Predictions: [66124.31387857934, 66229.10363492346, 61320.99895205716]
  - Feature Importance:

LSTM predictions (time-series forecasts):
  - Time series predictions: 36 total forecasts
  - Example predictions:
    * Profession 2374-9-005:
      Starting salary: 386448.00
      Salary after 12 months: 351968.00
      Change: -34480.00 (-8.92%)
    * Profession 2512-1-003:
      Starting salary: 419551.00
      Salary after 12 months: 401455.50
      Change: -18095.50 (-4.31%)
    * Profession 8322-9-005:
      Starting salary: 455123.00
      Salary after 12 months: 492049.00
      Change: 36926.00 (8.11%)
```

*Fig. 16: Integrated results from the Meta-Model*

## 4. Conclusion

This study presents a unified meta-model that integrates CatBoost, Explainable Boosting Machine (EBM), and LSTM neural networks with Bayesian optimization to address wage forecasting and labor demand prediction in Kazakhstan. By combining strengths in interpretability, prediction accuracy, and temporal forecasting, the meta-model offers a comprehensive tool for data-driven decision-making.

Its practical value is multi-faceted:

- Educational institutions can align curricula and training programs with labor market needs, maximizing student outcomes.
- Policymakers benefit from interpretable insights on wage disparities and saturation risks, enabling more effective policy intervention.
- Job seekers gain awareness of high-potential career paths based on forecasted wage trends and demand trajectories.
- Employers can plan compensation and hiring strategies based on real-time analytics of supply-demand mismatches.

While the model delivers strong results, especially with CatBoost achieving a MAPE of 2.51% and LSTM forecasting demand at 2.29% error, several limitations exist. Forecasting is currently limited to a 12-month horizon, and broader workforce variables such as job satisfaction or mobility are not yet integrated. These areas offer direction for future development.

Nonetheless, the proposed framework marks a significant advancement in labor market analytics, providing stakeholders with a reliable, interpretable, and scalable system to support strategic planning in an evolving economic landscape.

## References:

[1] Alshammari, R., Alshammari, T., Singh, S., & Rawat, D. B. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. Intelligent Systems with Applications, 18, 200228.

[2] Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological), 26(2), 211–243.

[3] Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1721–1730.

[4] Dorogush, A. V., Ershov, V., & Gulin, A. (2018). Catboost: Gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

[5] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

[6] Kaabar, S. (2024). Use catboost to predict your time series. Medium. Retrieved May 16, 2025.

[7] Karimov, A., Khasanov, T., & Ibragimov, R. (2025). Generalized meta framework for forecasting. ResearchGate.

[8] Liang, J., Jiang, L., Moreno, A., Zou, J., Fei-Fei, L., Niebles, J. C., & Sun, C. (2023). Advances in deep ensemble learning: Trends and perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(9), 10633–10651.

[9] Liu, J., Wang, H., Zhang, Y., & Chen, X. (2024). Two-stage meta-ensembling machine learning model for enhanced predictive analytics. Journal of Hydrology, 628, 130136.

[10] Microsoft Research. (2023). Train explainable boosting machines - classification. Microsoft Learn. Retrieved May 16, *2025.

[11] Nori, H., Jenkins, S., Koch, P., & Caruana, R. (2019). InterpretML: A unified framework for machine learning interpretability. arXiv preprint arXiv:1909.09223.

[12] Patel, R., Kumar, A., & Singh, V. (2024). LSTM adaptive hyperparameter tuning for financial time series forecasting. Journal of Theoretical and Applied Information Technology, 102(24).

[13] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31, 6638–6648.

[14] Ren, Y., Suganthan, P. N., & Srikanth, N. (2021). The ensemble approach to forecasting: A review and synthesis. Transportation Research Part C: Emerging Technologies, 132, 103357.

[15] Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. Advances in Neural Information Processing Systems, 25, 2951–2959.

[16] Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., & Deng, S.-H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. Journal of Electronic Science and Technology, 17(1), 26–40.

[17] Yeo, I.-K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. Biometrika, 87(4), 954–959.

[18] Zhang, W., Li, X., Chen, J., & Wang, Y. (2025). Developing an ensemble machine learning framework for enhanced predictive modeling. Nature Communications Earth & Environment, 6(1), 133.