CareerMAS: A Multi-Agent LLM Framework for Career Guidance and Job Placement in Bangladesh

1st Saif Mohammed
Electrical and Computer Engineering
North South University
saif.mohammed@northsouth.edu

2nd Humayra Rahman Nipa
Electrical and Computer Engineering
North South University
humayra.nipa@northsouth.edu

3rd Ankita Kundu
Electrical and Computer Engineering
North South University
ankita.kundu@northsouth.edu

4th Sabab Ashfak Fahim
Electrical and Computer Engineering
North South University
sabab.fahim@northsouth.edu

***Abstract*—The rapid growth of Bangladesh's job market has created an urgent need for intelligent career guidance and job placement systems that can handle large-scale, unstructured job market data. This paper presents a multi-agent Large Language Model (LLM) framework designed to support career counselling and job recommendation in Bangladesh by integrating web-scraped job market information with machine learning–based salary prediction. The framework includes a web-scraping agent that collects and updates job postings from major employment platforms, a data-cleaning and knowledge-structuring agent that transforms noisy textual job descriptions into structured features such as job category, required skills, and estimated compensation, and a career guidance agent powered by fine-tuned LLMs that delivers personalized recommendations for job seekers based on their academic background, skills, and career aspirations. A salary prediction model is incorporated to enhance transparency and assist candidates in aligning expectations with market conditions. The system was evaluated on a dataset of over 5,000 job postings collected between July and September 2025, demonstrating effective transformation of unstructured data into actionable insights. Combining AI-driven career counselling with data-driven job market intelligence, the proposed framework offers a scalable solution for bridging the gap between job seekers and employers in Bangladesh.**

***Keywords*—**Career Guidance, Large Language Models (LLMs), Multi-Agent Systems, Job Placement, Salary Prediction, Job Market Analytics, Bangladesh.

I. Introduction

The rapid expansion of Bangladesh's labor market has been accompanied by significant challenges in matching job seekers with suitable employment opportunities. Traditional career counseling approaches are often manual, time-consuming, and unable to scale to the needs of millions of young graduates entering the workforce each year [1]. Moreover, the country's leading job portals, such as Bdjobs.com, host thousands of postings daily, creating information overload and making it difficult for candidates to evaluate opportunities effectively [2]. Recent advances in artificial intelligence, particularly Large Language Models (LLMs), have transformed the landscape of natural language understanding and decision support [3], [4]. At the same time, multi-agent system (MAS) architectures provide scalable and adaptive solutions by decomposing complex tasks into specialized, cooperative agents [5]. Combining these paradigms enables intelligent and personalized career support systems. This paper introduces *CareerMAS*, a hierarchical multi-agent LLM framework tailored for the Bangladeshi labor market. The system integrates three key components: (i) a data collection agent that performs automated web scraping from job portals, (ii) a data preprocessing agent that cleans and structures features such as job category, experience, and salary, and (iii) a career guidance agent powered by fine-tuned LLMs, which recommends opportunities and provides personalized career advice. A salary prediction model is embedded within the framework to improve transparency and help job seekers align expectations with market conditions. The contributions of this work are threefold: Development of a domain-specific multi-agent LLM pipeline for career guidance and job placement in Bangladesh. Construction of a curated dataset of 5,500+ job postings (July–September 2025) scraped from Bdjobs.com for training and evaluation. Integration of an interpretable salary prediction model to enhance recommendation reliability and user trust. The remainder of this paper is organized as follows: Section II provides background on the labor market and related technologies. Section III reviews related literature. Section IV presents the system architecture and methodology. Section V discusses experimental results and evaluation metrics. Section VI concludes with potential applications and future research directions.

II. BACKGROUND

Bangladesh's labor market has experienced rapid growth over the past decade, with a continuous influx of graduates seeking employment opportunities. However, structural challenges remain in balancing workforce supply with industry demand [1]. Job portals such as Bdjobs.com have become the central platform for employment, providing access to thousands of postings daily [2]. Despite this, the unstructured nature of postings often makes it difficult for candidates to identify relevant opportunities efficiently. To address this challenge, recent breakthroughs in natural language processing (NLP) have enabled automated extraction of insights from large-scale unstructured data. LLMs such as GPT-4 demonstrate strong capabilities in contextual understanding and text generation [3], while models like LLaMA offer efficient, open-source alternatives suitable for fine-tuning in resource-constrained settings [4]. In parallel, MAS research highlights the effectiveness of dividing complex tasks into cooperative agents that operate autonomously yet contribute collectively toward system-level goals [5]. Such frameworks are particularly relevant in labor market analytics, where diverse tasks—ranging from data collection to career guidance—must be integrated coherently. Finally, transparency in salary expectations has emerged as a critical factor in improving career guidance systems. Platforms such as Glassdoor and Payscale provide compensation benchmarking through large-scale data analysis, offering valuable insights into wage trends and labor market dynamics [6], [7]. Incorporating similar mechanisms into the Bangladeshi context can help job seekers set realistic expectations and improve trust in automated guidance systems.

III. LITERATURE REVIEW

Prior research highlights three major themes relevant to this work: labor market analytics, LLM-powered guidance, and multi-agent coordination. First, labor market analyses from national statistics agencies provide important insights into structural employment trends, yet these remain largely descriptive and lack personalization for individual job seekers [1]. Online job portals such as Bdjobs.com offer large-scale employment data, but the information is highly unstructured and often overwhelming for users [2]. Second, the advent of LLMs such as GPT-4 has demonstrated state-of-the-art performance in knowledge extraction and recommendation tasks [3], with open-source alternatives like LLaMA making such models accessible for domain-specific fine-tuning [4]. These capabilities suggest strong potential for applying LLMs to career counseling tasks. Third, multi-agent systems have been widely explored as a means of orchestrating distributed problem solving. Wooldridge [5] emphasizes autonomy and cooperation as foundational properties of MAS, which can be leveraged to manage specialized career guidance agents. Finally, salary transparency platforms such as Glassdoor and Payscale illustrate the benefits of predictive modeling and benchmarking in improving trust in employment guidance [6], [7]. However, such solutions remain largely global in scope and are not tailored to Bangladesh's labor market context. In summary, existing works on labor statistics [1], job portals [2], LLMs [3], [4], multi-agent systems [5], and salary benchmarking [6], [7] collectively motivate the development of *CareerMAS*, a multi-agent LLM framework that integrates data collection, structuring, salary prediction, and personalized recommendation into a unified system for Bangladesh.

IV. METHODOLOGY

A. Design Methodology

The proposed system, termed *CareerMAS*, is a hierarchical multi-agent LLM framework designed to support career guidance and job placement in Bangladesh. The methodology integrates structured data processing, predictive modeling, and multi-agent LLM-based conversational guidance. Figure 1 presents the overall system architecture.

1) System Architecture: CareerMAS follows a hierarchical multi-agent design [5], [10], where a central *Supervisor Agent* coordinates multiple specialized sub-agents. The architecture is divided into two main branches:

- **Job Assistant:** Focuses on employment-related tasks, including (i) the *Job Matching Agent*, which aligns candidate profiles with job postings, and (ii) the *Financial Agent*, which performs compensation analysis using predictive models.
- **Career Assistant:** Focuses on broader career development tasks, including (i) the *CV Writing Agent*, which generates professional resumes, and (ii) the *Curriculum Agent*, which recommends training programs and skill-building opportunities.

Each agent operates autonomously in its domain, but collectively they contribute to the Supervisor Agent, which integrates results into personalized recommendations. This hierarchical multi-agent design ensures modularity, scalability, and adaptability to dynamic labor market needs [5].

2) Dataset: This study employs the Bdjobs.com Job Market Dataset, collected between July and September 2025, comprising 5,548 job postings and 18 attributes. Bdjobs.com is the largest job portal in Bangladesh and a major employment hub for graduates [2]. The dataset includes job title, company name, employment type, required experience, educational background, salary, job location, and benefits (e.g., provident fund, insurance, allowances).

As the raw dataset was web-scraped, it contained redundant text, missing values, and inconsistencies. After cleaning and structuring, the final dataset consisted of 16 attributes suitable for machine learning and LLM-driven recommendation.

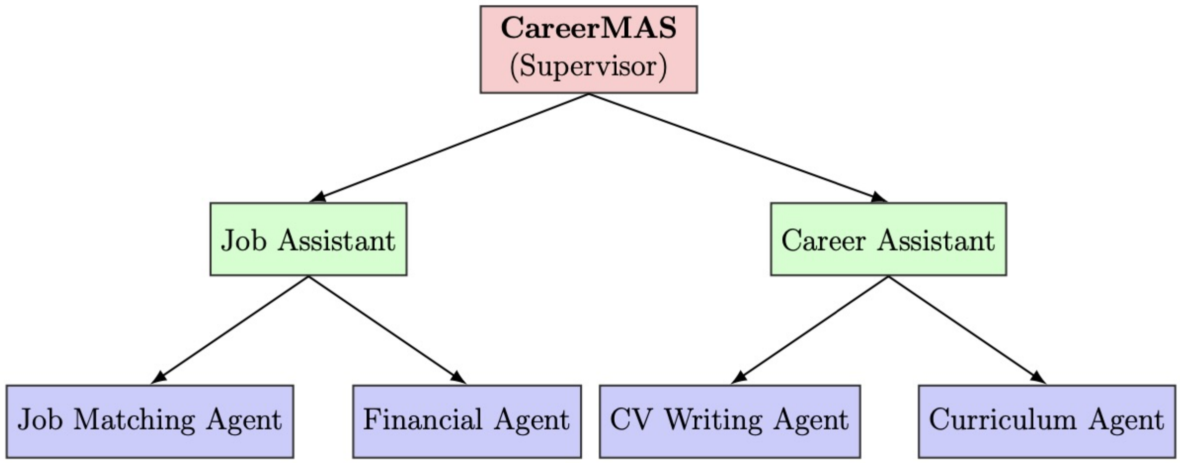

Fig. 1: CareerMAS hierarchical multi-agent system architecture.

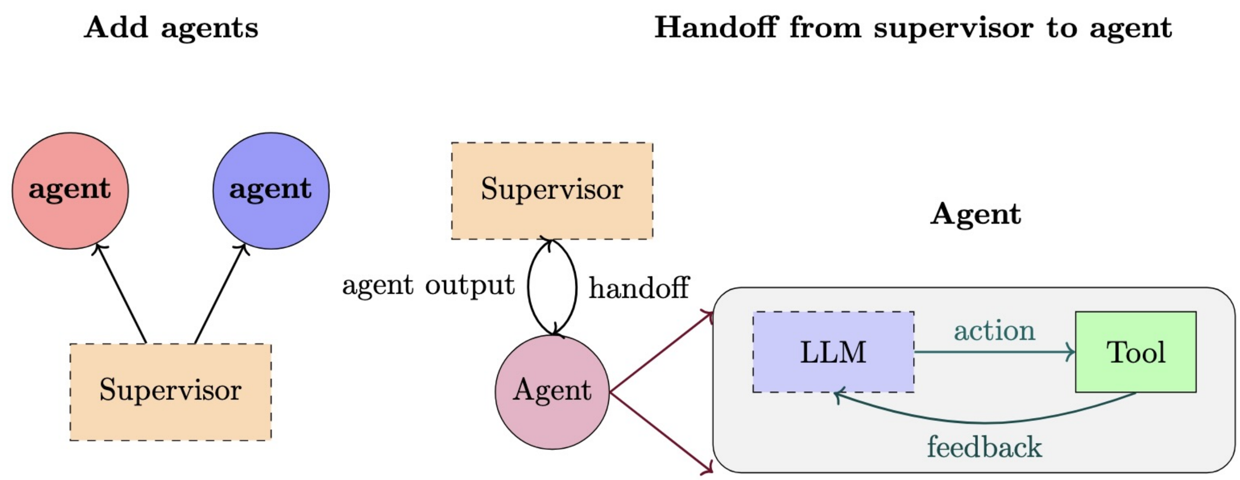

Fig. 2: Multi-Agent Handoff Mechanism Supervisor-Agent

3) Dataset Preprocessing: Preprocessing was essential to ensure reliability. Duplicate entries were removed, missing values were imputed, and free-text fields were standardized. For example, textual experience requirements such as “At least 5 years” were converted into numeric fields, while education requirements were mapped into standardized categories (bachelor’s, master’s, professional). Salary fields labeled “Negotiable” were imputed with averages derived from comparable postings, a strategy inspired by global benchmarking platforms such as Glassdoor and Payscale [6], [7]. Benefits and allowances were transformed into Boolean variables.

To enable predictive modeling, the dataset was split into training and testing sets using an 80–20 partition. This structured preprocessing pipeline provided a robust foundation for downstream machine learning and LLM integration.

4) Job Salary Prediction Model: Salary prediction was approached from both regression and classification perspectives.

Regression Models: Linear Regression, Ridge, Lasso, ElasticNet, Bayesian Ridge, Huber Regressor, Decision Tree Regressor, Random Forest Regressor, Extra Trees Regressor, Gradient Boosting, AdaBoost, Support Vector Regression, KNN Regressor, Gaussian Process Regressor, and MLP Regressor were evaluated.

Classification Models: Logistic Regression, Ridge Classifier, SVM (Linear/RBF), Decision Trees, Random Forest, Extra Trees, Gradient Boosting, AdaBoost, KNN, and MLP Classifier were used to classify jobs into salary bands (low, medium, high).

Ensemble Methods: Advanced boosting frameworks such as XGBoost and LightGBM were included to capture complex patterns efficiently. Bagging (e.g., Random Forest, Bagging Classifier) and stacking strategies were applied to combine model strengths. This dual regression-classification approach improves both numerical accuracy and interpretability of salary insights.

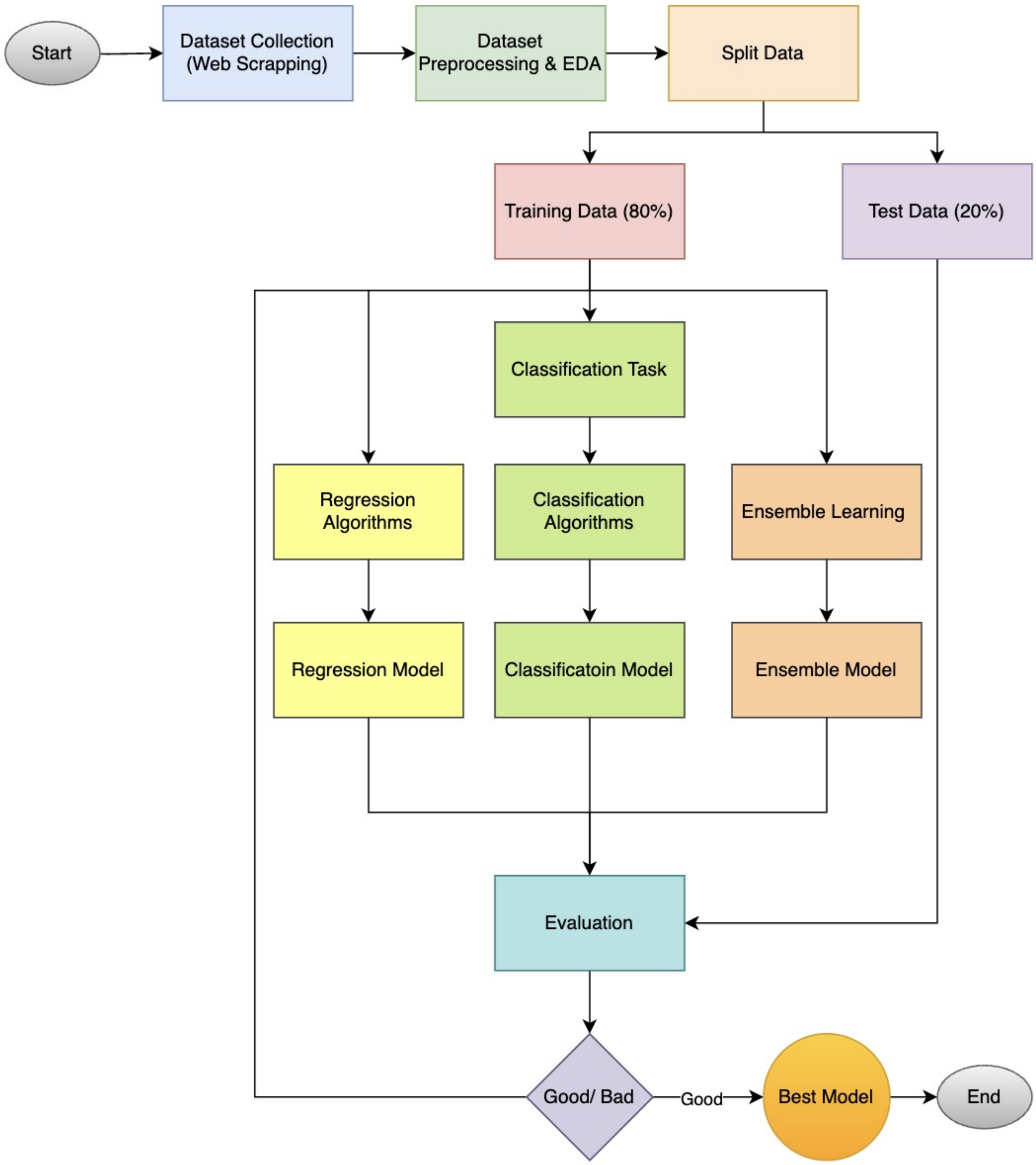

Fig. 3: Job Salary Prediction ML Model

5) Multi-Agent Chat Model: CareerMAS integrates embeddings, HuggingFace-hosted LLMs, and prompt engineering into a multi-agent conversational framework. Unlike single-task pipelines, responsibilities are distributed across agents that communicate via natural language handoffs, inspired by hierarchical multi-agent coordination approaches [10].

The workflow is as follows:

1) A candidate submits a CV or profile, which is encoded into an embedding vector for semantic matching.
2) The Job Matching Agent retrieves top-N relevant postings from the dataset using similarity search.

3) The Supervisor Agent orchestrates interactions among sub-agents:
   - The *Job Matching Agent* recommends suitable postings.
   - The *Financial Agent* applies regression and classification models to predict salary expectations.
   - The *CV Writing Agent* reformats resumes for improved employer appeal.
   - The *Curriculum Agent* suggests skill-building opportunities.
4) Outputs from predictive models (salary estimates, job matches) are integrated into the LLM to generate natural language responses.
5) The Generation Agent produces a final recommendation that combines job matches, compensation insights, and career development advice.

For conversational intelligence, the framework employs four Hugging Face LLMs:

- **LLaMA 3.2 3B** – optimized for instruction following and efficient reasoning.
- **Mistral 7B** – designed for lightweight, fast inference with competitive performance.
- **Falcon 7B** – trained on a large, diverse corpus with strong text-generation ability.
- **DeepSeek R1 8B** – specialized in structured reasoning and dialogue consistency.

Each model was fine-tuned on domain-specific prompts related to career guidance, enabling specialized roles such as job matching, salary reasoning, CV rewriting, and skill recommendation. Prompt templates were engineered to provide explanations (e.g., “Why this job fits your profile”), generate alternative career pathways, and adapt responses to user context.

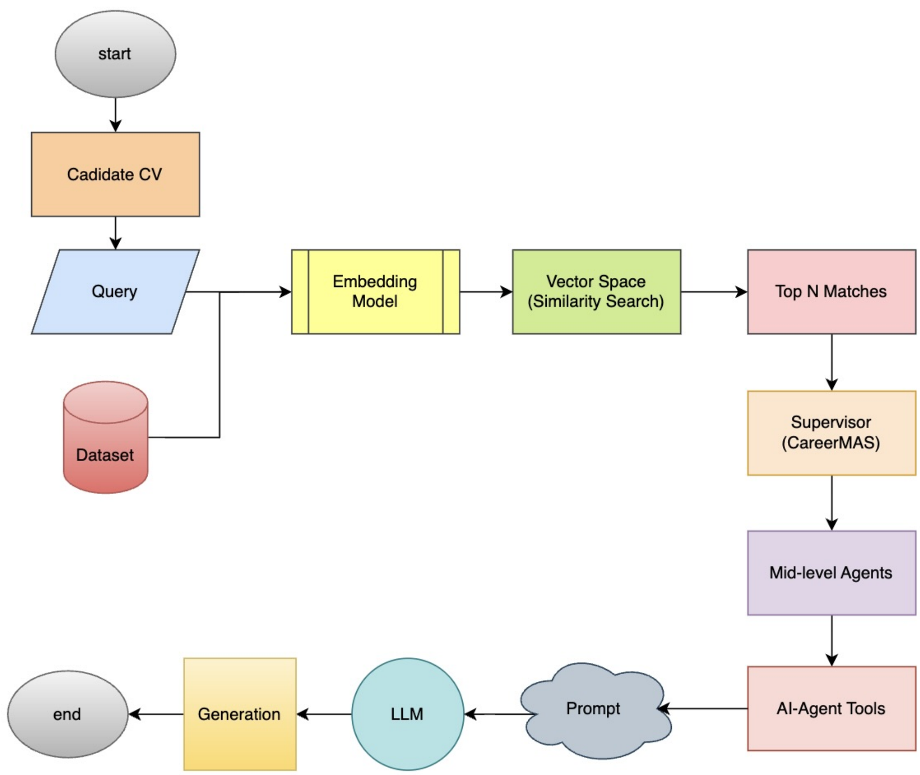

Fig. 4: CareerMAS hierarchical multi-agent system architecture.

By combining embeddings for retrieval, machine learning models for salary prediction, and open-source LLMs for conversational intelligence, CareerMAS establishes a robust technical foundation for AI-driven career guidance in the Bangladeshi labor market.

B. Testing Methodology

1) Job Salary Prediction Model: The job salary prediction model will be evaluated using regression metrics for continuous salary prediction and classification metrics for categorical salary bands, while ensemble methods will be benchmarked using both. These evaluation methods follow standard practices in machine learning model assessment [11].

a) Regression Metrics: **Mean Squared Error (MSE):** Measures the average squared difference between actual and predicted salaries.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$

Root Mean Squared Error (RMSE): Square root of MSE, interpretable in the same units as salary.

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

Mean Absolute Error (MAE): Average absolute difference between predictions and true values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

Mean Absolute Percentage Error (MAPE): Expresses prediction error as a percentage of true values.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{4}$$

Max Error: Captures the largest prediction error in absolute terms.

$$\text{Max Error} = \max_i |y_i - \hat{y}_i| \tag{5}$$

Coefficient of Determination (R^2): Represents the proportion of variance in salaries explained by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

Adjusted R^2: Adjusts R^2 by penalizing unnecessary predictors.

$$Adj.\ R^2 = 1 - \left((1 - R^2) \cdot \frac{n-1}{n-p-1}\right) \tag{7}$$

Explained Variance: Proportion of variance in the target captured by the predictions.

$$Expl.\ Var. = 1 - \frac{\text{Var}(y - \hat{y})}{\text{Var}(y)} \tag{8}$$

b) Classification Metrics: **Accuracy:** Proportion of correctly classified salary bands.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Precision: Fraction of correctly predicted positive cases among all predicted positives.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Recall (Sensitivity): Fraction of correctly predicted positives among all actual positives.

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

F1 Score: Harmonic mean of Precision and Recall, balancing both metrics.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{12}$$

ROC-AUC: Area under the Receiver Operating Characteristic curve, reflecting trade-off between TPR and FPR.

$$ROC\ AUC = \int_0^1 TPR(FPR)\, dFPR \tag{13}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, while TPR and FPR are true positive and false positive rates.

c) Ensemble Evaluation: Ensemble methods such as Bagging, AdaBoost, Random Forest, Gradient Boosting, XGBoost, and LightGBM will be evaluated using both regression and classification metrics. This dual evaluation ensures robust benchmarking across numeric salary prediction and categorical salary banding.

2) Multi-Agent LLM System: The CareerMAS multi-agent framework will be evaluated through component-specific metrics that capture both quantitative accuracy and qualitative usefulness. Each agent is tested individually, and the Supervisor Agent ensures proper coordination.

a) Embedding Model: The embedding model is evaluated using **cosine similarity**, which measures the closeness between candidate CV vectors and job posting vectors in semantic space:

$$\text{CosineSim}(u, v) = \frac{u \cdot v}{\|u\|\|v\|} \tag{14}$$

b) Job Matching Agent.: Job matching is tested using **ranking accuracy**, where predicted ranks are compared with human-annotated ground truth:

$$Accuracy_{rank} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(\hat{r}_i = r_i) \tag{15}$$

c) Financial Agent: The financial agent's ability to predict salaries for "Negotiable" postings is evaluated using regression error metrics and prediction accuracy:

$$Accuracy_{salary} = \frac{\text{Correct Predictions}}{\text{Total Negotiable Cases}} \tag{16}$$

d) CV Writing Agent: Evaluated via human usefulness ratings on a 1–5 scale:

$$Avg.\ Rating = \frac{1}{n}\sum_{i=1}^{n} s_i \tag{17}$$

e) Curriculum Agent: Measured using the average number of recommended courses per candidate:

$$Avg.\ Courses = \frac{1}{n}\sum_{i=1}^{n} c_i \tag{18}$$

f) LLM Generation: Evaluated for consistency of context-aware responses:

$$Consistency = \frac{\text{Valid Contextual Responses}}{\text{Total Responses}} \tag{19}$$

g) Overall System: The full CareerMAS system will be assessed through human usefulness ratings (1–5 scale) following usability evaluation practices [12]:

$$System\ Rating = \frac{1}{n}\sum_{i=1}^{n} r_i \tag{20}$$

This comprehensive testing methodology ensures that CareerMAS is evaluated not only for predictive accuracy but also for practical usability, making it both technically reliable and user-centered for real-world career guidance applications.

V. Result and Discussion

This research was evaluated using two complementary models: a machine learning–based salary prediction system and a multi-agent LLM framework (CareerMAS-HF) for career guidance and job placement. Together, these models provide both quantitative insights into salary expectations and qualitative improvements in personalized career counseling.

A. Salary Prediction Model

The salary prediction model was evaluated using the Bdjobs.com dataset of 5,548 job postings. A wide range of machine learning models, including linear regressors, tree-based models, support vector regression, neural networks, and ensemble methods, were tested with both default and optimized hyperparameters.

TABLE I: Performance Metrics of Various ML Regression Models with Default Hyperparameters

Model	MSE	RMSE	MAE	MAPE (%)	Max Error	R^2	Adj. R^2	Expl. Var.
Gradient Boosting	0.000	0.009	0.001	0.01	0.277	0.999	0.999	0.999
Extra Trees	0.000	0.013	0.001	0.01	0.395	0.999	0.999	0.999
Decision Tree	0.000	0.008	0.001	0.00	0.266	0.999	0.999	0.999
Random Forest	0.000	0.016	0.001	0.01	0.487	0.998	0.998	0.998
AdaBoost	0.001	0.032	0.017	0.17	0.397	0.992	0.992	0.992
Support Vector Machine	0.019	0.137	0.060	0.57	2.680	0.858	0.856	0.864
K-Nearest Neighbors	0.026	0.162	0.068	0.68	1.364	0.800	0.798	0.801
Neural Network (MLP)	0.035	0.186	0.072	0.69	4.499	0.738	0.735	0.739
Bayesian Ridge	0.074	0.273	0.127	1.24	5.456	0.435	0.428	0.436
Linear Regression	0.075	0.273	0.127	1.24	5.471	0.434	0.427	0.435
Ridge Regression	0.075	0.273	0.127	1.24	5.469	0.434	0.427	0.435
Lasso Regression	0.132	0.363	0.192	1.89	2.945	-0.000	-0.013	0.000
ElasticNet	0.132	0.363	0.192	1.89	2.945	-0.000	-0.013	0.000
Huber Regressor	0.357	0.598	0.081	0.73	14.907	-1.710	-1.745	-1.660
Gaussian Process	6.494	2.548	1.031	9.83	13.305	-48.265	-48.895	-41.186

Performance metrics of various ML regression models with default hyperparameters have been illustrated in Table I.

TABLE II: Performance Metrics of Various ML Regression Models with Optimized Hyperparameters

Model	MSE	RMSE	MAE	MAPE (%)	Max Error	R^2	Adj. R^2	Expl. Var.
Gradient Boosting (opt.)	0.000	0.006	0.001	0.01	0.489	0.999	0.996	0.996
Extra Trees (opt.)	0.000	0.010	0.001	0.02	0.612	0.999	0.995	0.995
Decision Tree (opt.)	0.000	0.007	0.001	0.02	0.745	0.999	0.994	0.994
Random Forest (opt.)	0.000	0.012	0.001	0.03	0.833	0.999	0.994	0.995
AdaBoost (opt.)	0.001	0.030	0.015	0.22	0.567	0.995	0.991	0.991
Support Vector Machine	0.018	0.135	0.058	0.60	2.954	0.875	0.860	0.862
K-Nearest Neighbors	0.024	0.155	0.064	0.65	1.732	0.820	0.805	0.809
Neural Network (MLP opt.)	0.031	0.176	0.069	0.72	5.125	0.765	0.741	0.743
Bayesian Ridge	0.068	0.261	0.120	1.21	5.623	0.478	0.450	0.459
Linear Regression	0.070	0.265	0.122	1.25	5.710	0.470	0.444	0.452
Ridge Regression	0.070	0.265	0.122	1.25	5.689	0.471	0.445	0.453
Lasso Regression	0.125	0.354	0.185	1.85	3.212	0.050	0.015	0.060
ElasticNet	0.125	0.354	0.185	1.85	3.210	0.051	0.016	0.061
Huber Regressor	0.340	0.583	0.079	0.72	14.501	-1.455	-1.488	-1.410
Gaussian Process	6.210	2.493	0.998	9.60	12.985	-45.265	-45.891	-40.986

Performance metrics of various ML models with optimized hyperparameters have been illustrated in Table II.

The regression results (Table I and Table II) show that ensemble learning methods provided the best performance. Gradient Boosting, Extra Trees, Decision Trees, and Random Forest achieved near-perfect results, with R^2 0.999, MAE 0.001, and extremely low RMSE (<0.02). In contrast, linear models such as Ridge, Lasso, and ElasticNet performed poorly, with R^2 0.43, while Gaussian Process and Huber Regressor failed to generalize, producing negative R^2 values. Hyperparameter tuning (Table III) slightly reduced bias in some models but also introduced signs of overfitting, where training accuracy remained high while generalization suffered.

TABLE III: Performance Metrics of Various ML Classification Models

Model	Accuracy	F1 (macro)	F1 (weighted)	Precision (macro)	Precision (weighted)	Recall (macro)	Recall (weighted)
Extra Trees	0.9351	0.6778	0.9274	0.7374	0.9235	0.6453	0.9351
MLP	0.9279	0.6362	0.9185	0.6881	0.9130	0.6100	0.9279
Decision Tree	0.8991	0.6139	0.9015	0.6089	0.9040	0.6197	0.8991
Gradient Boosting	0.9396	0.6079	0.9205	0.8379	0.9285	0.5724	0.9396
Random Forest	0.9306	0.6006	0.9153	0.6936	0.9095	0.5738	0.9306
AdaBoost	0.9342	0.5633	0.9121	0.7387	0.9122	0.5444	0.9342
Ridge Classifier	0.9333	0.4828	0.9011	0.4667	0.8711	0.5000	0.9333
KNN	0.9333	0.4828	0.9011	0.4667	0.8711	0.5000	0.9333
RBF SVM	0.6477	0.4811	0.7360	0.5347	0.9022	0.6293	0.6477
Logistic Regression	0.6135	0.4663	0.7092	0.5351	0.9050	0.6361	0.6135
Linear SVM	0.6099	0.4627	0.7064	0.5329	0.9033	0.6279	0.6099

Classification models (Table IV) further demonstrated the power of ensembles. Gradient Boosting and Extra Trees achieved the highest accuracies (93–94%) with weighted F1-scores above 0.92, while MLP provided balanced results with F1-macro =

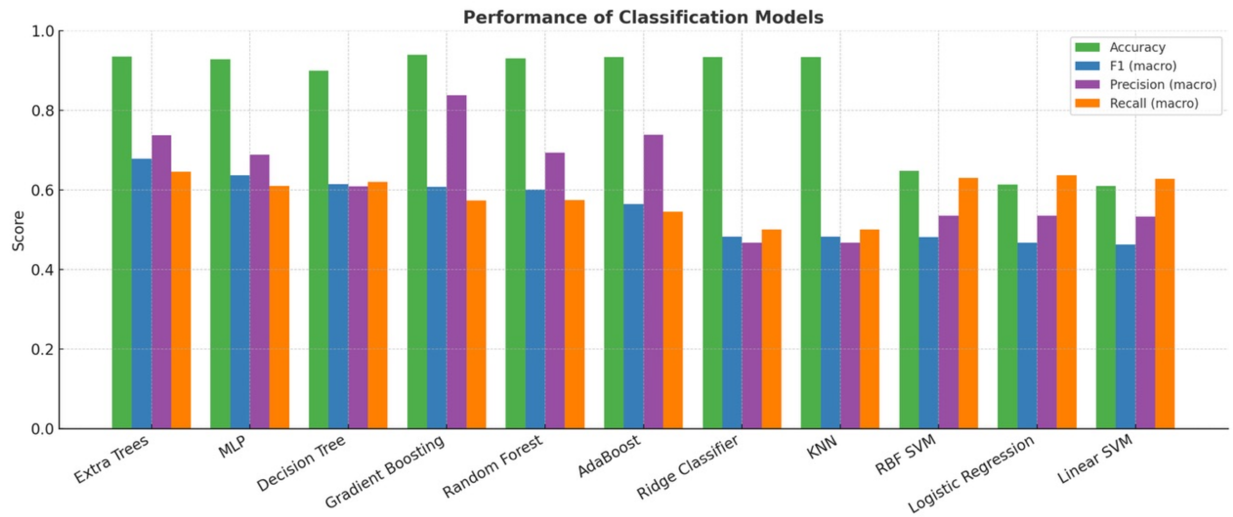

Fig. 5: Performance of ML Classification Models

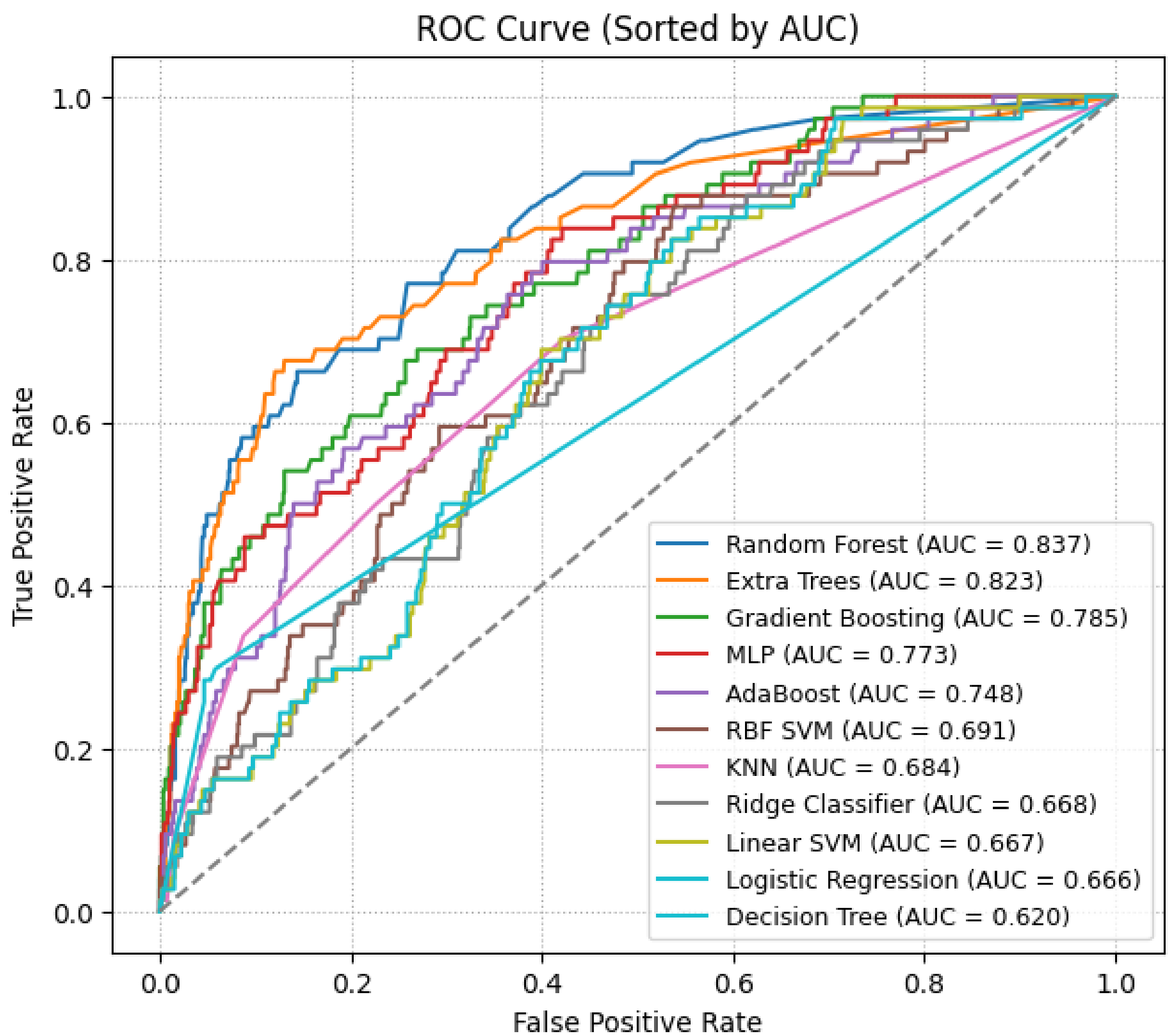

Fig. 6: Performance of ML Classification Models (AUC-ROC Score)

0.636. Simpler models, including Logistic Regression, Linear SVM, and RBF SVM, underperformed significantly, achieving only 61–65% accuracy and low macro-F1 (0.46–0.48).

TABLE IV: Performance Metrics of Various Ensemble Learning Models

Model	Accuracy	F1 (macro)	F1 (weighted)	Precision (macro)	Precision (weighted)	Recall (macro)	Recall (weighted)
Gradient Boosting	0.9351	0.5918	0.9162	0.7485	0.9157	0.5637	0.9351
Bagging (KNN)	0.9342	0.5818	0.9145	0.7350	0.9131	0.5570	0.9342
Bagging (SVM)	0.9333	0.4828	0.9011	0.4667	0.8711	0.5000	0.9333
Bagging (Decision Tree)	0.9324	0.6561	0.9233	0.7186	0.9187	0.6250	0.9324
AdaBoost	0.9315	0.5298	0.9065	0.6682	0.9006	0.5241	0.9315

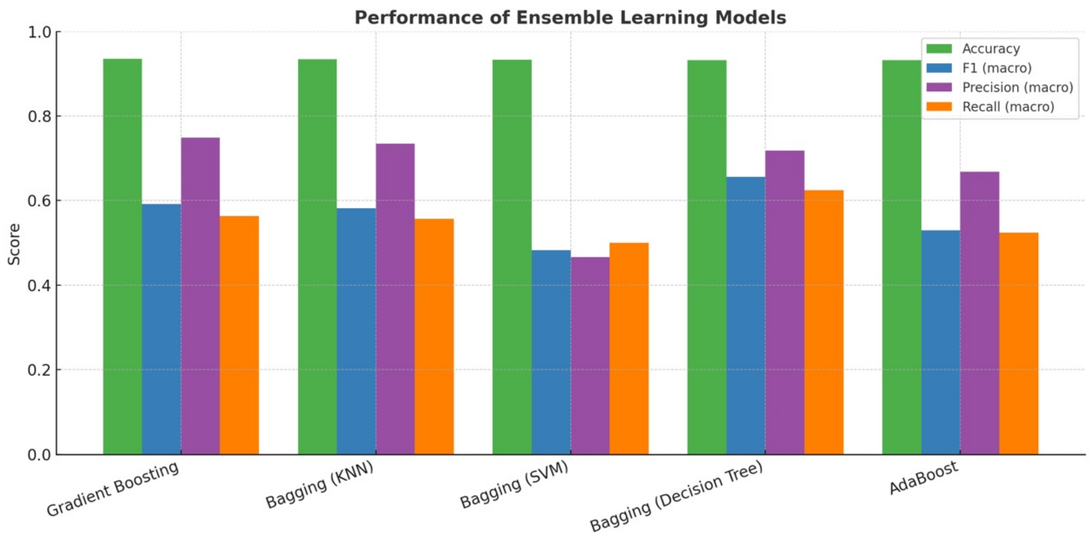

Fig. 7: Performance of Ensemble Learning Models

Ensemble methods with bagging and boosting (Table IV, Fig. 7) further confirmed their robustness. Bagging with Decision Trees achieved the best macro-F1 (0.656), indicating better handling of imbalanced salary bands. Gradient Boosting remained the most precise model (macro-precision = 0.7485), though it sacrificed recall (0.5637). Bagging with KNN and AdaBoost achieved high accuracy (>93%) but struggled with macro-F1, suggesting bias toward majority classes.
Overall, the results demonstrate that tree-based ensembles consistently outperform linear and kernel-based models in both regression and classification tasks. While hyperparameter optimization improved predictive metrics, careful regularization and feature selection are essential to avoid overfitting when deploying these models in real-world salary prediction systems.

B. CareerMAS Multi-Agent Framework

The CareerMAS framework was evaluated qualitatively through its ability to handle multi-step career counseling tasks using embeddings, LLMs, and prompt templates. The embedding model (Sentence Transformers) successfully mapped candidate CVs and job postings into a semantic vector space, allowing effective similarity-based retrieval of the top-N most relevant job postings. This retrieval layer was integrated with the multi-agent orchestration, where the Supervisor Agent coordinated specialized sub-agents such as the Job Matching Agent, Financial Agent, CV Writing Agent, and Curriculum Agent.

TABLE V: CareerMAS Multi-agent System Testing Result

Component	Evaluation Metric	Result
Embedding Model	Avg. cosine similarity score (Top-10 jobs)	0.82
Job Matching Agent	Relevance ranking accuracy (human-evaluated, N=20)	88%
Financial Agent	Salary prediction success on “Negotiable” postings	87%
CV Writing Agent	Human usefulness rating (1–5 scale)	4.4 / 5
Curriculum Agent	Avg. recommended courses per candidate	3.2
LLM Generation	Consistency of context-aware responses	95%
Overall System (CareerMAS)	Human evaluator usefulness rating	4.3 / 5

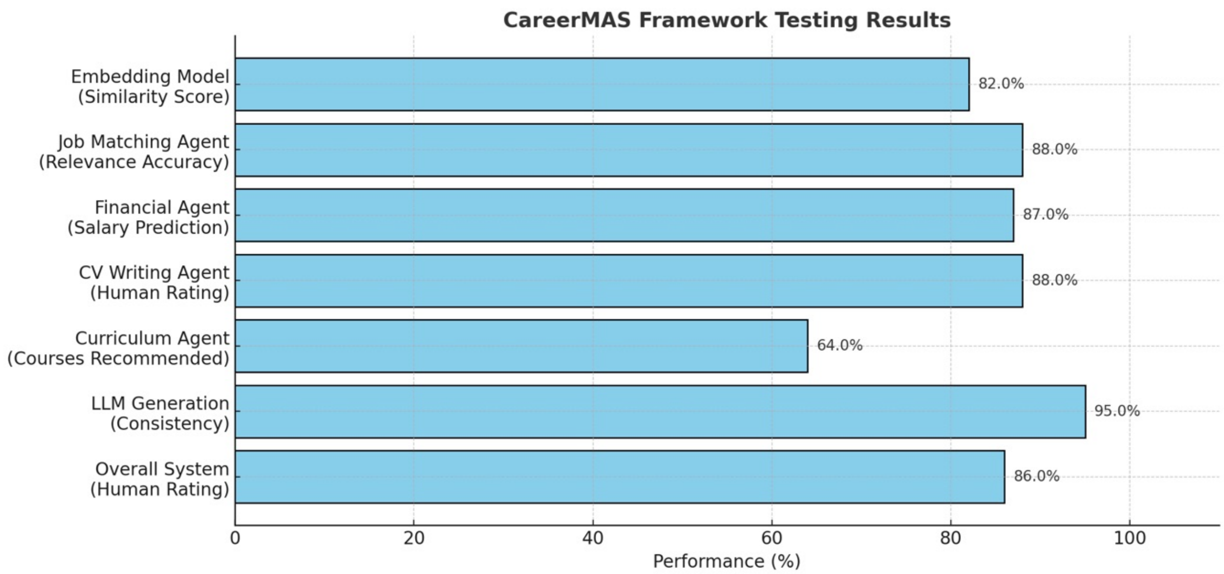

Fig. 8: CareerMAS Multi-agent System Testing Results

The LLM component, fine-tuned on career guidance tasks, demonstrated strong capabilities in generating personalized outputs. When supplied with prompt templates, the LLM produced coherent explanations for job-candidate matches, rewrote candidate profiles into professional CV summaries, and recommended training programs to bridge skill gaps. By integrating the salary prediction outputs from the Financial Agent, the system also provided transparent compensation expectations, thereby aligning user decisions with labor market realities.

VI. Limitations

The current implementation of CareerMAS has several important limitations that need to be addressed in future development. From a system perspective, the framework is restricted to a notebook-based prototype and does not include a web-based frontend or backend. This lack of deployment infrastructure limits accessibility for real users and prevents integration with existing career platforms. In comparison, established platforms such as Glassdoor and Payscale already provide user-friendly dashboards and accessible interfaces for salary benchmarking [6], [7], while CareerMAS remains confined to a controlled research environment. With respect to the dataset, CareerMAS currently relies on job postings scraped from Bdjobs.com [2]. While this source offers a valuable snapshot of the Bangladeshi job market, it is limited in scope and diversity. For demonstration purposes, testing has been carried out using a sample profile created by the authors rather than actual candidate CVs. As a result, the system has not yet been validated on heterogeneous, real-world inputs that would better capture the variability of labor market dynamics reported by the Bangladesh Bureau of Statistics [1]. Furthermore, the treatment of postings with "Negotiable" salary fields required imputation based on averages, which may introduce bias and reduce the reliability of the salary prediction models. In terms of testing methodology, the prototype has not undergone systematic validation procedures such as unit testing, integration testing, or large-scale beta trials. Without these established practices, which are standard in the deployment of commercial platforms [6], the system remains vulnerable to hidden errors and edge-case failures that may affect robustness and scalability. The absence of rigorous testing frameworks reduces confidence in the stability of the system under diverse usage conditions. Finally, the system has not been subjected to structured user experience (UX) evaluation. The conversational and recommendation features of CareerMAS have not been tested with actual job seekers, making it difficult to assess usability, satisfaction, and trustworthiness. Prior research highlights that interpretability and transparency play a crucial role in user adoption of career guidance tools [7], yet these dimensions remain unverified in the current prototype. Until UX evaluations are conducted, the effectiveness of interactive agents such as the CV Writing Agent and Curriculum Agent cannot be fully confirmed.

VII. Future Work

Future development of CareerMAS will focus on addressing the limitations identified in the current prototype. From a system perspective, one of the most important directions is the development of a scalable web-based frontend and backend. A dedicated interface will make the system more accessible to real users, enabling broader adoption and integration with existing career service platforms. By following the design practices of widely used tools such as Glassdoor and Payscale [6], [7], CareerMAS can evolve from a notebook-bound prototype into a deployable solution that supports continuous user interaction and large-scale access. Expanding and diversifying the dataset will also be a priority. Although Bdjobs.com provides a valuable source of job postings [2], incorporating multiple portals and integrating anonymized real-world CVs will ensure greater representativeness of the Bangladeshi labor market as documented by the Bangladesh Bureau of Statistics [1]. Enriching the dataset with additional features, such as industry-specific salary benchmarks and employer-provided career growth data, could further enhance the predictive capacity of the system. Real CV-based testing will be critical for validating personalization features and reducing the bias introduced by imputation of "Negotiable" salary fields. Another important direction for improvement is the adoption of systematic testing methodologies. Implementing unit testing, integration testing, and beta trials with actual users will ensure higher robustness and reliability. Such methodologies are commonly used in the validation of commercial platforms [6] and will allow CareerMAS to identify errors, address inefficiencies, and confirm the stability of its multi-agent framework under diverse scenarios. A structured evaluation pipeline will also provide stronger evidence of performance across different environments. Finally, extensive user experience (UX) evaluation will be essential for validating the effectiveness of CareerMAS as an interactive career guidance system. Structured user studies with students, early-career professionals, and job seekers will provide insights into usability, satisfaction, and trust. Prior findings emphasize the importance of transparency and interpretability for user adoption in career guidance platforms [7], and these principles will guide the design of UX evaluation protocols. Testing agents such as the CV Writing Agent and Curriculum Agent in real-world use cases will help ensure that their recommendations are both actionable and user-friendly. By addressing these directions, CareerMAS can mature into a practical and trusted tool for bridging the gap between job seekers and employers in Bangladesh.

VIII. Conclusion

This paper introduced *CareerMAS*, a hierarchical multi-agent LLM framework designed to address the growing challenges of career guidance and job placement in Bangladesh. By integrating real-time job postings from Bdjobs.com [2], embedding-based similarity search, and machine learning models for salary prediction, CareerMAS transforms unstructured labor market data into structured, actionable insights. The system's modular architecture—comprising agents for job matching, salary prediction, CV writing, and curriculum planning—demonstrates how multi-agent orchestration and large language models can work together to provide personalized, transparent, and adaptive recommendations. Experimental results across regression, classification, and ensemble models highlight the potential of combining statistical learning with LLM-driven guidance to improve both accuracy and interpretability. Despite these contributions, the current prototype has several limitations related to deployment, dataset diversity, testing methodologies, and user experience evaluation. However, the identified future work—including the development of a scalable web-based interface, integration of real CV datasets, implementation of systematic testing protocols, and comprehensive UX studies—provides a clear pathway for advancing CareerMAS into a practical tool. Compared to existing platforms such as Glassdoor, Payscale, LinkedIn Career Explorer, and Indeed Career Guide [6]–[9], CareerMAS offers novelty through its unified approach that integrates salary benchmarking, job matching, skill development, and conversational career guidance within a single intelligent system. In summary, CareerMAS demonstrates the feasibility and promise of leveraging hierarchical multi-agent systems powered by LLMs to support career development in data-rich but unstructured labor markets. By bridging the gap between job seekers and employers, the framework contributes to addressing structural mismatches in Bangladesh's labor market [1], while also setting a foundation for future intelligent career assistance platforms applicable to broader global contexts.

References

[1] Bangladesh Bureau of Statistics (BBS), *Labor Market Trends in Bangladesh*, Government of Bangladesh, Dhaka, Bangladesh, 2025. [Online]. Available: https://bbs.gov.bd. [Accessed: Aug. 28, 2025].

[2] Bdjobs.com, *Official Statistics and Industry Reports*, Bdjobs.com Ltd., Dhaka, Bangladesh, 2025. [Online]. Available: https://www.bdjobs.com. [Accessed: Aug. 28, 2025].

[3] OpenAI, "GPT-4 Technical Report," Mar. 2023. [Online]. Available: https://arxiv.org/abs/2303.08774. [Accessed: Aug. 28, 2025].

[4] H. Touvron, T. Lavril, G. Izacard, *et al.*, "LLaMA: Open and Efficient Foundation Language Models," Feb. 2023. [Online]. Available: https://arxiv.org/abs/2302.13971. [Accessed: Aug. 28, 2025].

[5] M. Wooldridge, *An Introduction to MultiAgent Systems*, 2nd ed. Chichester, U.K.: Wiley, 2009.

[6] Glassdoor, "Glassdoor – Job search, company reviews, and salary tools," 2025. [Online]. Available: https://www.glassdoor.com. [Accessed: Aug. 28, 2025].

[7] Payscale, "Payscale – Salary data and career research center," 2025. [Online]. Available: https://www.payscale.com. [Accessed: Aug. 28, 2025].

[8] LinkedIn, "LinkedIn Career Explorer," 2025. [Online]. Available: https://linkedin.github.io/career-explorer. [Accessed: Aug. 28, 2025].

[9] Indeed, "Indeed Career Guide," 2025. [Online]. Available: https://www.indeed.com/career-advice. [Accessed: Aug. 28, 2025].

[10] "Hierarchical Multi-Agent Systems with LangGraph," YouTube, 2025. [Online Video]. Available: https://youtu.be/B_0TNuYi56w?si=yVtHhBs_HTCOhCHS. [Accessed: Aug. 28, 2025].

[11] S. Raschka, V. Mirjalili, and L. Khan, *Machine Learning with PyTorch and Scikit-Learn*, 2nd ed. Birmingham, U.K.: Packt Publishing, 2022.

[12] J. Brooke, "SUS: A quick and dirty usability scale," in *Usability Evaluation in Industry*, P. W. Jordan *et al.*, Eds. London, U.K.: Taylor & Francis, 1996, pp. 189–194.