





Intelligent Assistant for Resources Optimization and information retrieval

Eng. Amjad Homedan & Eng. Anas Alhabashi

A Thesis Submitted to

Faculty of Computers and Information Technology

Fulfillment of the Requirements for the Master Degree of Computer

Science

Under the Supervision of

Prof. Dr. Yasser Abdelhamid

Associate Professor, Faculty of Computers and information

Midocean University

Content

1

Introduction

2

Problem definition and Objectives

3

Integration Between Resources Optimization and inelegance Assistant

4

Statistical Analysis in Resource Optimization

5

Thesis Methodologies used

6

Experimental Results and Model review

7

Conclusion and Future work

Introduction

This thesis explores the integration of machine learning and AI-driven solutions for resource optimization with intelligent Q&A Assistant. It develops a framework that leverages predictive modeling techniques, such as Recursive Feature Elimination (RFE) and advanced algorithms like Random Forest and Gradient Boosting, to enhance resource allocation and intelligence Q&A assistant. The proposed solution addresses key operational challenges through automation of repetitive tasks, better decision-making, and alignment of human resources to project requirements. This work will further outline the synergies between Human Resource Optimization and Document Intelligence to improve productivity and reduce inefficiencies, thus enabling organizations to meet their goals efficiently in today's competitive environment.

Abstract

The AI-driven document summarization system uses AI-based technologies to extract, analyze, and summarize information from uploaded DOCX files with the use of Flask. It processes uploaded files for key details and returns data in JSON format for easy integration with other systems. It will make document processing easier, support project management, and improve decision-making with its summarized but accurate data. It automates document extraction and summarization, facilitating reporting, reducing manual efforts, and ensuring data accuracy, hence becoming an indispensable tool in efficient workflow optimization.

AI-Powered Developer Prediction Model is a machine learning-based solution for predicting the required number of backend and frontend developers in web development projects. Using Random Forest models integrated with a Flask API, it allows for real-time predictions based on the analysis of project size, type, duration, and cost. This system allows for better resource planning, optimized project delivery timelines, and more effective cost control by automating what was previously an explicit task of estimating the number of developers required. Accurately and efficiently allocating resources reduces operational risks, enhances better decision-making, and ensures the successful execution of projects with cost efficiency.

AI-Powered PDF Content Extraction and Q&A Model This is an advanced AI-powered solution for content extraction, chunking, and analysis of PDFs to perform query-based tasks. By using tools like pdfminer, FAISS, and OpenAI embeddings, the system efficiently processes large document repositories for fast and accurate answers to user queries. It enhances business intelligence, reduces manual workload, and improves operational efficiency by providing precise and context-aware responses. This intelligent PDF query system provides knowledge management and automated customer support, smoothing access to the most valuable information and making the experience much more friendly for users.

Benefits of Proposed Framework

Optimizes resource allocation in real-time.

Enhances Q&A Assistant adaptability and accuracy.

Summarized Project document to assist highlighting the main Inputs.

Improves operational efficiency and decision-making.

Machine learning empowers smarter resource allocation.

Q&A Assistant become more efficient and error-free.

1

Introduction

2

Problem definition and Objectives

3

Integration Between Resources Optimization and inelegance Assistant

4

Statistical Analysis in Resource Optimization

5

Thesis Methodologies used

6

Experimental Results and Model review

7

Conclusion and Future work

Problem definition and Objectives

Resource optimization and intelligent Q&A assistants are key enablers for organizations to pursue their goals of efficiency, better decision-making, and improved customer experiences. Traditional methods lack the capability to manage complex resources, predict future demands, and provide timely and accurate responses to user queries. Data-driven solutions, on the other hand, apply advanced analytics and AI to analyze historical data, identify patterns, and make informed predictions. These systems excel in key areas such as resource allocation, where they fit the right resources to tasks based on skills, availability, and priorities while anticipating future needs and addressing potential bottlenecks. This also extends to task management, where they prioritize, schedule, and track tasks, optimize workflows, and drive productivity. It is also further enhanced with the operations of intelligent Q&A assistants that provide accurate and relevant answers, learning from the history of interactions to understand and serve users' needs better. By adopting data-driven approaches, organizations can considerably improve operational efficiency, reduce costs, and adapt to the evolving technological demands-irreplaceable for the success of contemporary business.

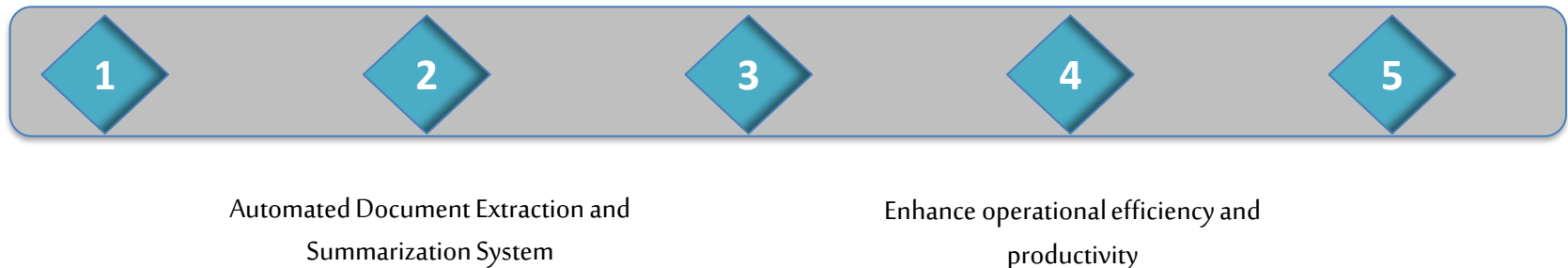


Main Objectives

Predict future resource needs and address bottlenecks proactively

Provide accurate, timely, and relevant responses using intelligent Q&A Assistant

Reduce costs and improve customer satisfaction



- 1 Introduction
- 2 Problem definition and Objectives
- 3 **Integration Between Resources Optimization and inelegance Assistant**
- 4 Statistical Analysis in Resource Optimization
- 5 Thesis Methodologies used
- 6 Experimental Results and Model review
- 7 Conclusion and Future work

Integration Between Resources Optimization and Q&A inelegance Assistant

By creates a powerful synergy that helps organizations create efficiency in operations and better decision-making. Resource optimization, through deep AI and machine learning, addresses the efficient allocation of available human, technical, and financial resources to meet identified priorities of the organization. At the same time, intelligent Q&A assistants help improve productivity by automating information gathering, answering users' queries faster and more accurately, and reducing repetitive work. These put together empower businesses to forecast the demand for resources, take proactive obstruction, and provide actionable insights. This will not only help in smoothing out the operational efficiencies but also supports informed decision-making, facilitates agility, and improves customer satisfaction by correctly aligning resources to real-time organizational needs.

Main Principals



Develop predictive models for resource optimization



Enhance Q&A Model through machine learning and NLP.



Use Statistical Modeling like MSE, RMSE, R2, Log-Likelihood for evaluation

1

Introduction

2

Problem definition and Objectives

3

Integration Between Resources Optimization and inelegance Assistant

4

Statistical Analysis in Resource Optimization

5

Thesis Methodologies used

6

Experimental Results and Model review

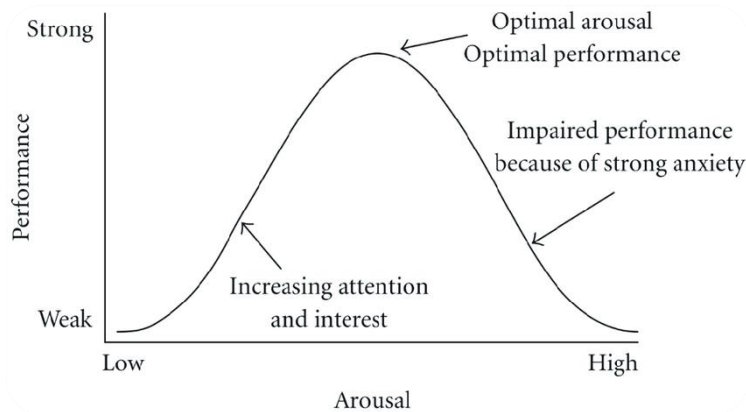
7

Conclusion and Future work

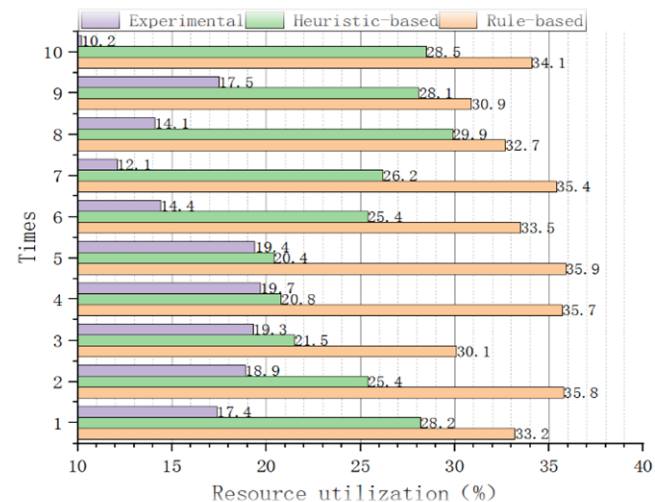
Most of the reference consider the below 3 Principles in their Research

- 1 Uses event logs to understand resource behavior and improve allocation strategies.
- 2 Focuses on performance metrics to inform resource allocation decisions.
- 3 Includes insights on workload and its impact on performance.

Statistical Examples



Hebbian version of the Yerkes-Dodson law



(Li & Zhang, 2024)

Content

1

Introduction

2

Problem definition and Objectives

3

Integration Between Resources Optimization and inelegance Assistant

4

Statistical Analysis in Resource Optimization

5

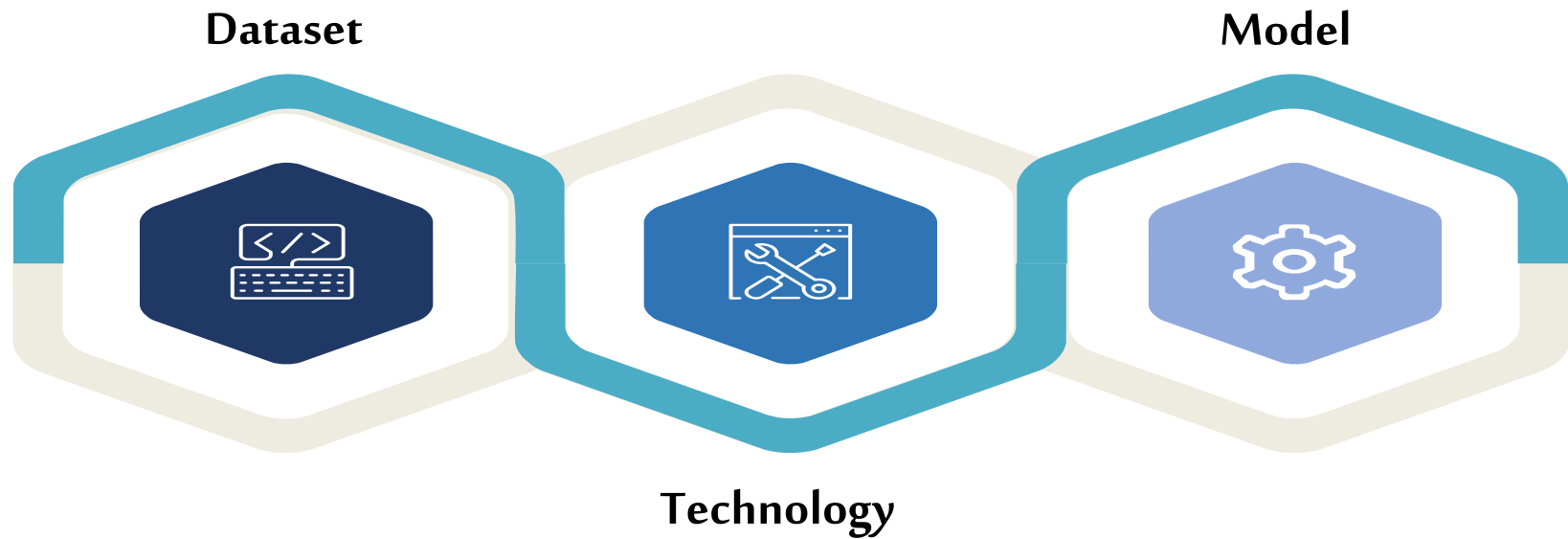
Thesis Methodologies used

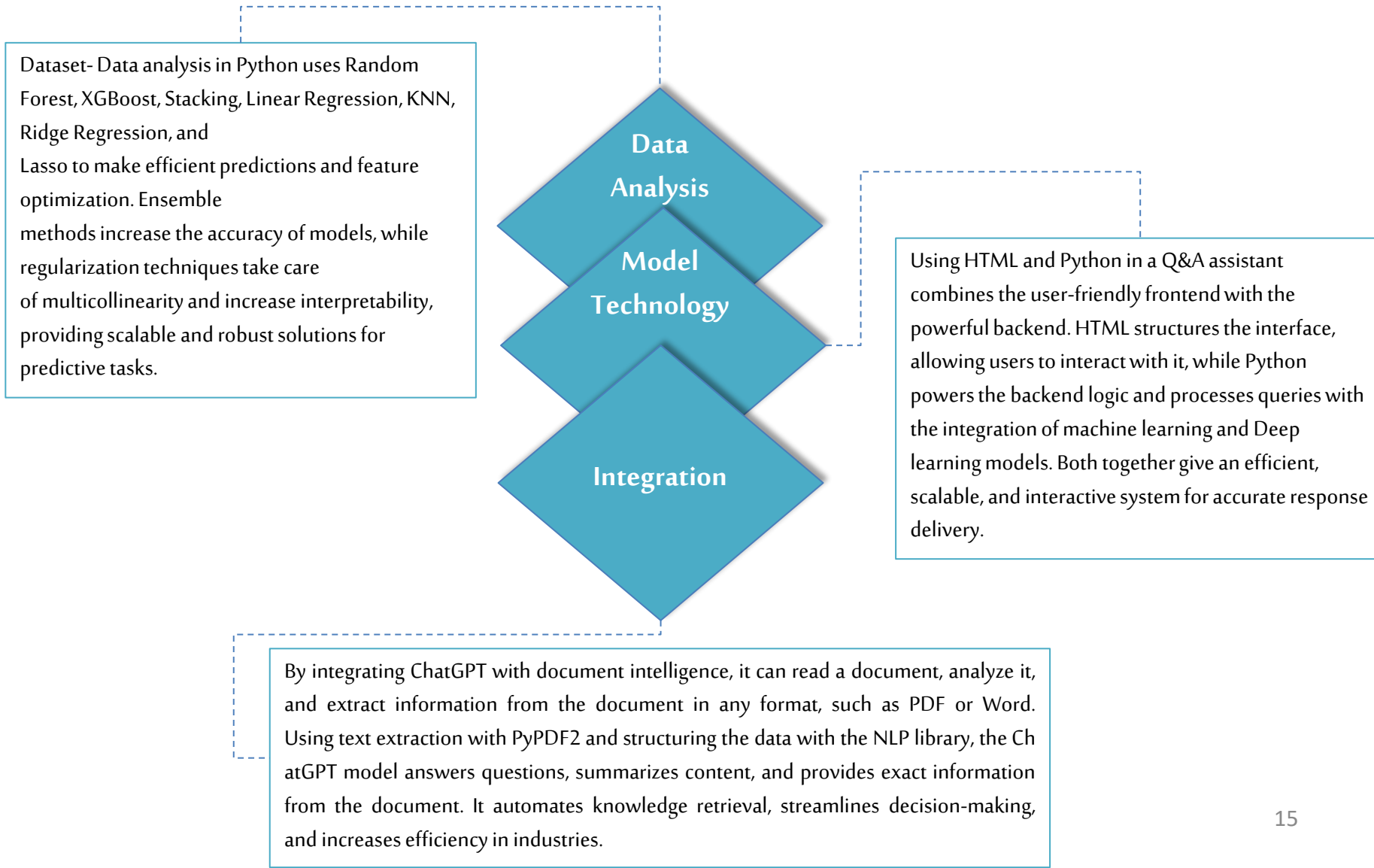
6

Experimental Results and Model review

7

Conclusion and Future work





We have prepared a dataset with a range of attributes and further extended this with the inclusion of numeric coding 0/1, with regard to project size to capture scale, hence becoming better at predicting the project cost and resource requirement estimates. The dataset has been split into 75% training and 25% test sets to perform some good model evaluation. Project category was added to describe project urgency and importance; it will provide an input in the next modeling with priorities.

Sample Dataset

project name	project_price	project_duration_ month	project_size_sma ll	project_size_me dium	project_size_large	project_type_re gular	project_type_vip	backend_developer_cou nt	frontend_developer_count
Social Media Platform	1000000	6	1	0	0	0	1	2	3
Membership Platform	1005000	6	1	0	0	0	1	2	1
Membership Platform	1005022	6	1	0	0	0	1	2	3
Event Management Site	1005045	6	1	0	0	0	1	2	3
Community Forum	1005080	6	1	0	0	0	1	2	1
Portfolio Showcase	1006000	6	1	0	0	0	1	2	1
Event Management Site	1008000	6	1	0	0	0	1	2	3

Content

1

Introduction

2

Problem definition and Objectives

3

Integration Between Resources Optimization and inelegance Assistant

4

Statistical Analysis in Resource Optimization

5

Thesis Methodologies used

6

Experimental Results and Model review

7

Conclusion and Future work



Intelligent Assistance In Resource Plan

Project Summary

Upload Document: proje...o.docx

Project Name: : Project_F

Project Price: : \$450,000

Client Name: : Client_F

Project Employee Resources Required:

: System Analyst, Business Analyst,
Backend Developer

Project Duration (months): : 18

Project Description Summary: : This project aims to develop a comprehensive web application for a client in the financial

Predict Backend and Frontend

Project Size: ▼

Project Price:

project duration:

Project Type: ▼

Expected Backend Developers: 3

Expected Frontend Developers: 2

Ask AI ChatBot about Project

project_details_pmo.pdf

PDF uploaded and processed successfully

Answer: The project price is \$450,000.

1

PDF and DOCX file processing.

2

Predictive modeling using machine learning (Random Forest Regressor).

3

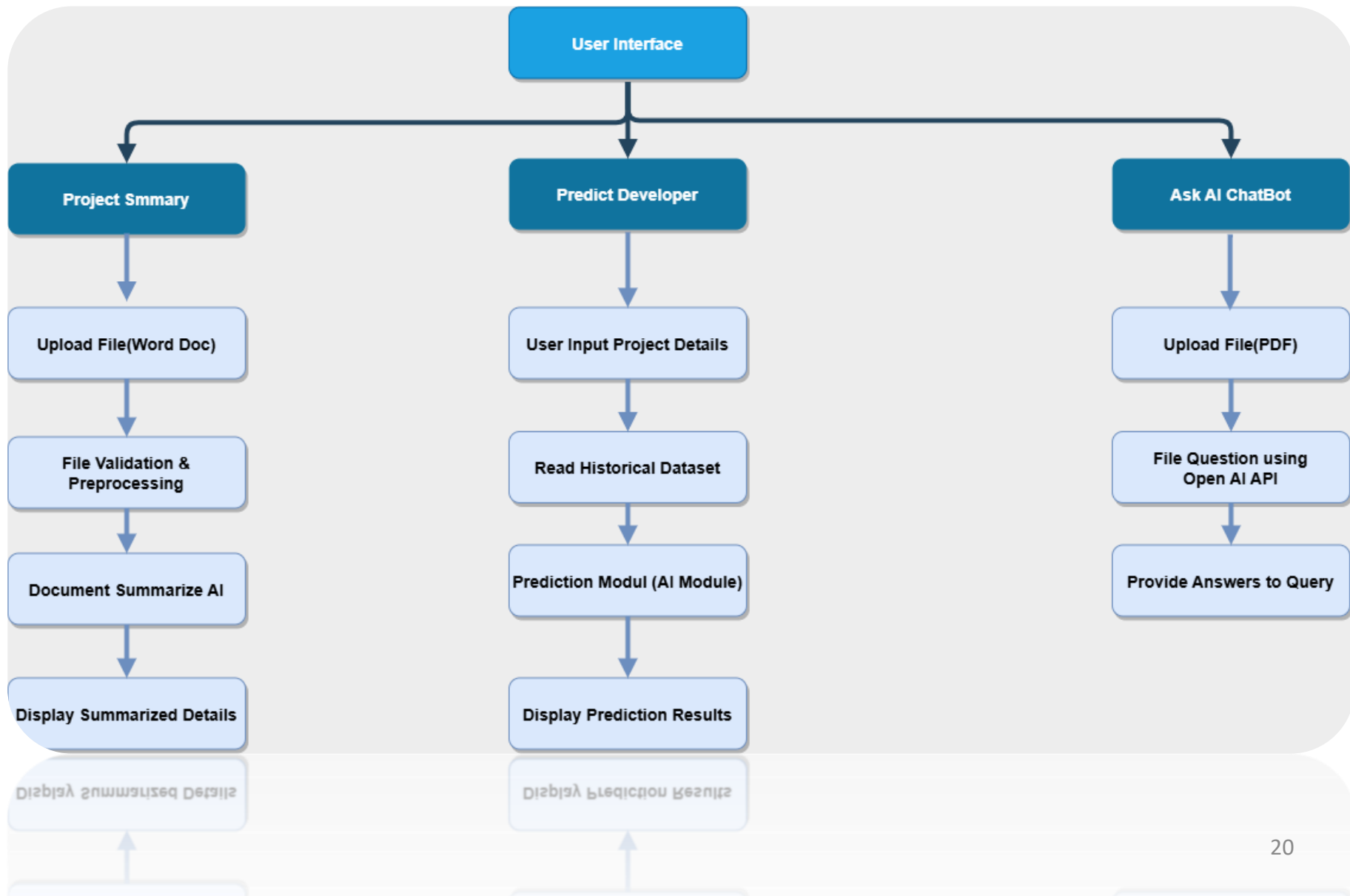
Backend developer count prediction based on project attributes.

4

Question-answering using pre-processed text chunks.

5

Multiple endpoints for user interaction, file uploads, and predictions



Machine Learning Prediction

Predict the number of front and backend developers required for a project.

1

Preprocess the dataset
(Website_Development
_Project_Dataset.csv

2

Train a Random Forest
Regressor using
Pipeline with standard
scaling

3

Expose the model via
/predict API to accept
project details and
return predictions

1

Flask: For creating the web application and handling HTTP requests.

2

Pandas: For data manipulation.

3

Scikit-learn: For data preprocessing and predictive modeling.

4

Python-docx: For processing .docx files.

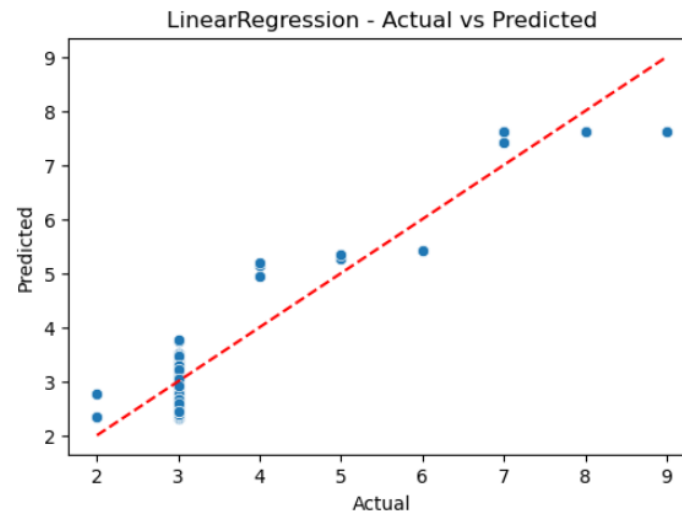
5

Werkzeug: For secure file handling.

Function	Description
Project Document extraction	Extract and Summarize the important information within the Word document.
Extract_pdf_text(uploaded_file)	Extracts text content from a PDF file using pdfminer. Logs and handles errors during extraction.
Split_text(text, max_length)	Splits the input text into smaller chunks, with a maximum length per chunk for efficient processing.
Create_vector_store(chunks)	Creates a semantic vector store using FAISS and embeds text chunks with OpenAI's OpenAIEmbeddings.
Num_tokens(text)	Counts the number of tokens in the given text using tiktoken.
Retrieve_relevant_chunks(...)	Retrieves the most relevant chunks from the vector store based on a question, ensuring token limits.
Ask_question_openai(...)	Queries OpenAI's GPT-4 model with the relevant text and a question to generate a context-based answer.

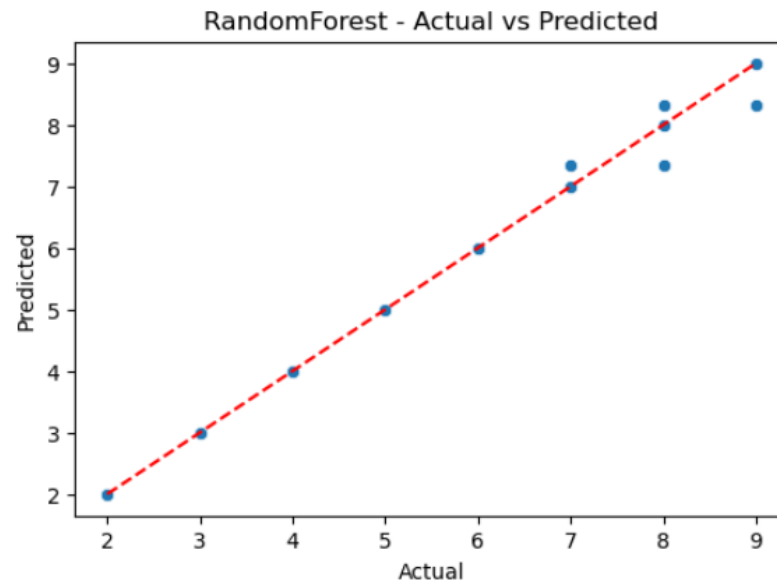
1 Linear Regression

The evaluation metrics for the Linear Regression model show a Mean Squared Error (MSE) of 0.268145, indicating a moderate average squared difference between actual and predicted values. The Root Mean Squared Error (RMSE) is 0.517827, representing the error magnitude in the same units as the target variable. The model has an R^2 value of 0.877454, suggesting that approximately 87.75% of the variance in the target variable is explained by the model, which reflects a good fit. The Log-Likelihood is -2239.108340, which is relatively negative, but higher values would be preferable. Lastly, the Cost, based on a custom cost function, is 0.046020, indicating the model's predictive Acc with a relatively low error. Overall, the model performs well with a high R^2 and reasonable error and cost metrics.



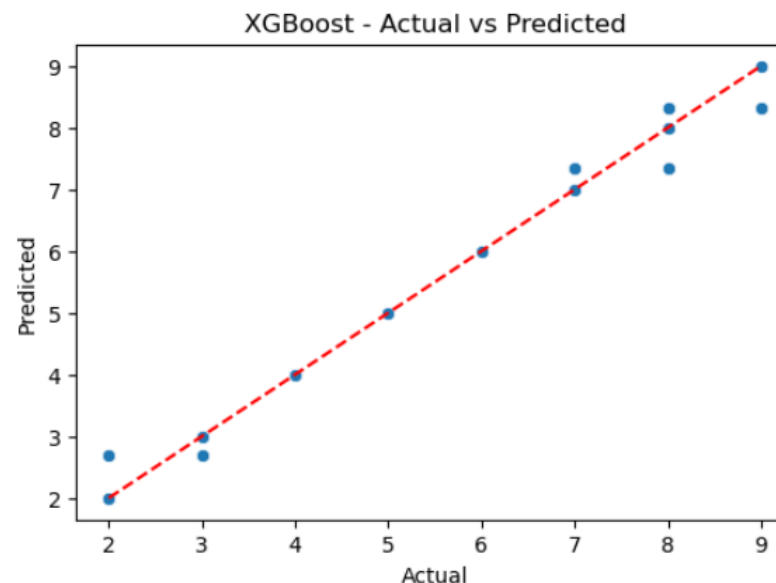
Random Forest

The evaluation metrics for the Random Forest model reveal a Mean Squared Error (MSE) of 0.023846, indicating a relatively small average squared difference between the actual and predicted values, which suggests that the model performs well. The Root Mean Squared Error (RMSE) is 0.154421, which is low, indicating a good fit with minimal error. The R^2 value is 0.989102, meaning the model explains approximately 98.91% of the variance in the target variable, which is a very strong performance. The Log-Likelihood is 1321.792596, showing that the model fits the data well in terms of probability distribution. Finally, the Cost, calculated using a custom cost function, is 0.004093, which is extremely low, further indicating the model's effectiveness. Overall, the Random Forest model shows excellent predictive performance with high R^2 , low error, and minimal cost.



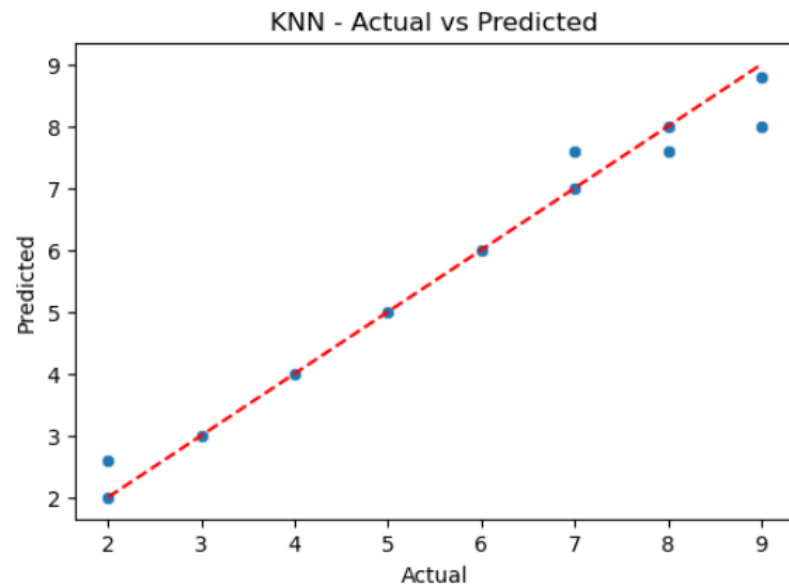
3 XGBoost

The evaluation metrics for the XGBoost model show a Mean Squared Error (MSE) of 0.024904, indicating a small average squared difference between the actual and predicted values, similar to Random Forest. The Root Mean Squared Error (RMSE) is 0.157809, which is slightly higher than Random Forest but still suggests good predictive Acc. The R^2 value is 0.988619, meaning the model explains about 98.86% of the variance in the target variable, indicating a very strong performance. The Log-Likelihood is 1257.933056, reflecting a good fit with the data in terms of probability distribution. Lastly, the Cost is 0.004274, which is still low, confirming that the model performs well with minimal error. Overall, XGBoost shows excellent performance with high R^2 , low error, and minimal cost.



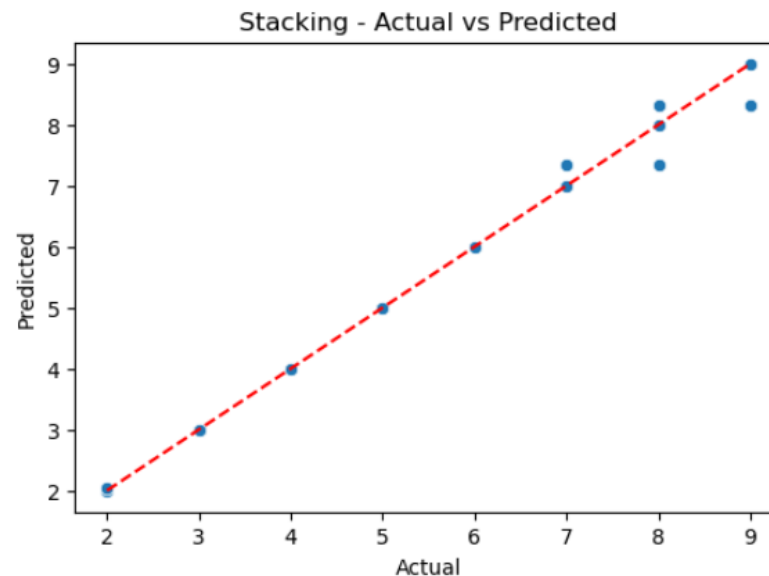
The evaluation metrics for the K-Nearest Neighbors (KNN) model show a Mean Squared Error (MSE) of 0.031152, which is slightly higher compared to XGBoost and Random Forest. The Root Mean Squared Error (RMSE) is 0.176499, indicating a reasonable level of prediction error, though higher than the other models. The R^2 value of 0.985763 suggests that KNN explains about 98.58% of the variance in the target variable, which is still a strong performance. The Log-Likelihood of 928.521545 reflects the model's fit in terms of probability distribution. The Cost is 0.005346, which is slightly higher compared to XGBoost but still relatively low, suggesting that the KNN model is performing well with manageable error. Overall, KNN provides good results but slightly lags behind in terms of MSE and RMSE compared to the other models.

k-Nearest Neighbor (KNN)



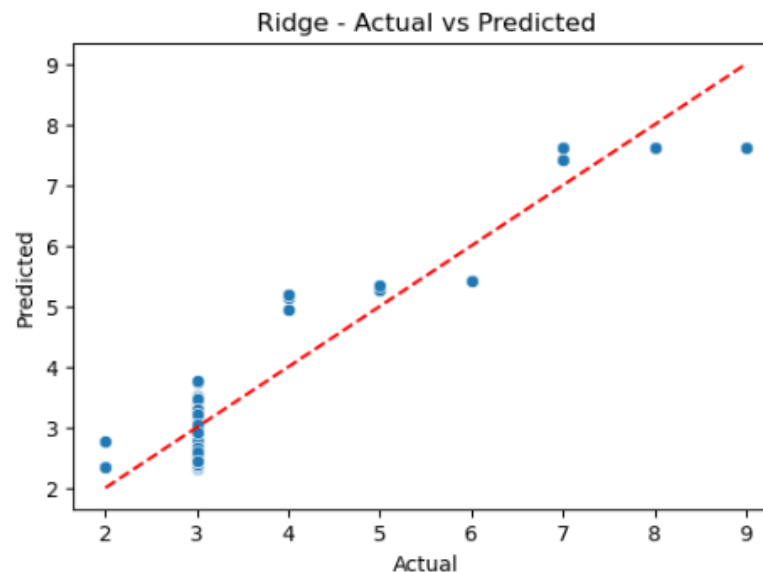
The Stacking model evaluation metrics show an MSE of 0.023846, which is very close to that of the Random Forest model, indicating low prediction error. The RMSE value of 0.154421 also suggests good Acc, with a relatively small deviation between actual and predicted values. The R^2 value of 0.989102 indicates that the model explains about 98.91% of the variance in the target variable, which is excellent. The Log-Likelihood value of 1321.806576 demonstrates a strong model fit based on the residuals' distribution. The Cost is 0.004092, which is among the lowest compared to other models, indicating good efficiency in minimizing prediction error relative to the true values. Overall, the Stacking model performs exceptionally well, combining the strengths of multiple base models for superior prediction Acc.

Stacking



The Ridge model evaluation metrics indicate an MSE of 0.268145, which is relatively higher compared to other models, suggesting a moderate level of prediction error. The RMSE value of 0.517827 reflects a noticeable deviation between actual and predicted values. The R^2 value of 0.877454 indicates that the model explains approximately 87.75% of the variance in the target variable, which is still a good fit but not as high as the Stacking or Random Forest models. The Log-Likelihood of -2239.106535 is significantly lower, which suggests that the model may not be as well-fit as the other models, and the residuals are not perfectly aligned with the assumed distribution. The Cost of 0.046020 is the highest among the models, indicating relatively higher inefficiency in minimizing error. In summary, while Ridge performs reasonably well, it does not outperform models like Stacking or Random Forest in terms of Acc and cost efficiency.

Ridge Regression



Models Comparison and contrast

Model	MSE	RMSE	R ²	Cost
LinearRegression	0.223354	0.472604	0.897924	0.038333
GradientBoosting	0.051377	0.226666	0.97652	0.008818
RandomForest	0.051408	0.226732	0.976506	0.008823
XGBoost	0.051377	0.226665	0.97652	0.008818
KNN	0.05597	0.23658	0.974421	0.009606
Ridge	0.223354	0.472604	0.897924	0.038333
Lasso	0.224886	0.474221	0.897224	0.038596
Stacking	0.052016	0.22807	0.976228	0.008927

1	Introduction
2	Problem definition and Objectives
3	Integration Between Resources Optimization and inelegance Assistant
4	Statistical Analysis in Resource Optimization
5	Thesis Methodologies used
6	Experimental Results and Model review
7	Conclusion and Future work

Conclusion and Key Takeaways



Machine learning greatly optimizes resources by using models such as Random Forest, Stacking, XGBoost, KNN, Ridge, and Lasso. Of these, Random Forest and Stacking showed the best results with low MSE and RMSE, which can be interpreted by high accuracy of prediction and robust minimization of error. The ensemble models were particularly effective in explaining up to 87% of the variance in resource-related data, hence optimizing their allocation and reducing computational costs.



In Resource Management and intelligent Q&A Assistant, it helps in efficient allocations of human, material, and time resources by predicting demands, optimizing schedules, and improving task assignments. It enhances decision-making through automation in error detection, forecasting, and performance evaluation in intelligent Q&A assistants. Such a system provides real-time feedback, detects inefficiencies, and proactively predicts issues with the least downtime.



ML not only optimizes resource allocation but also strengthens intelligent Q&A Assistant, enabling businesses to reduce costs, improve efficiency, and make smarter decisions through predictive modeling.

Future Work Directions



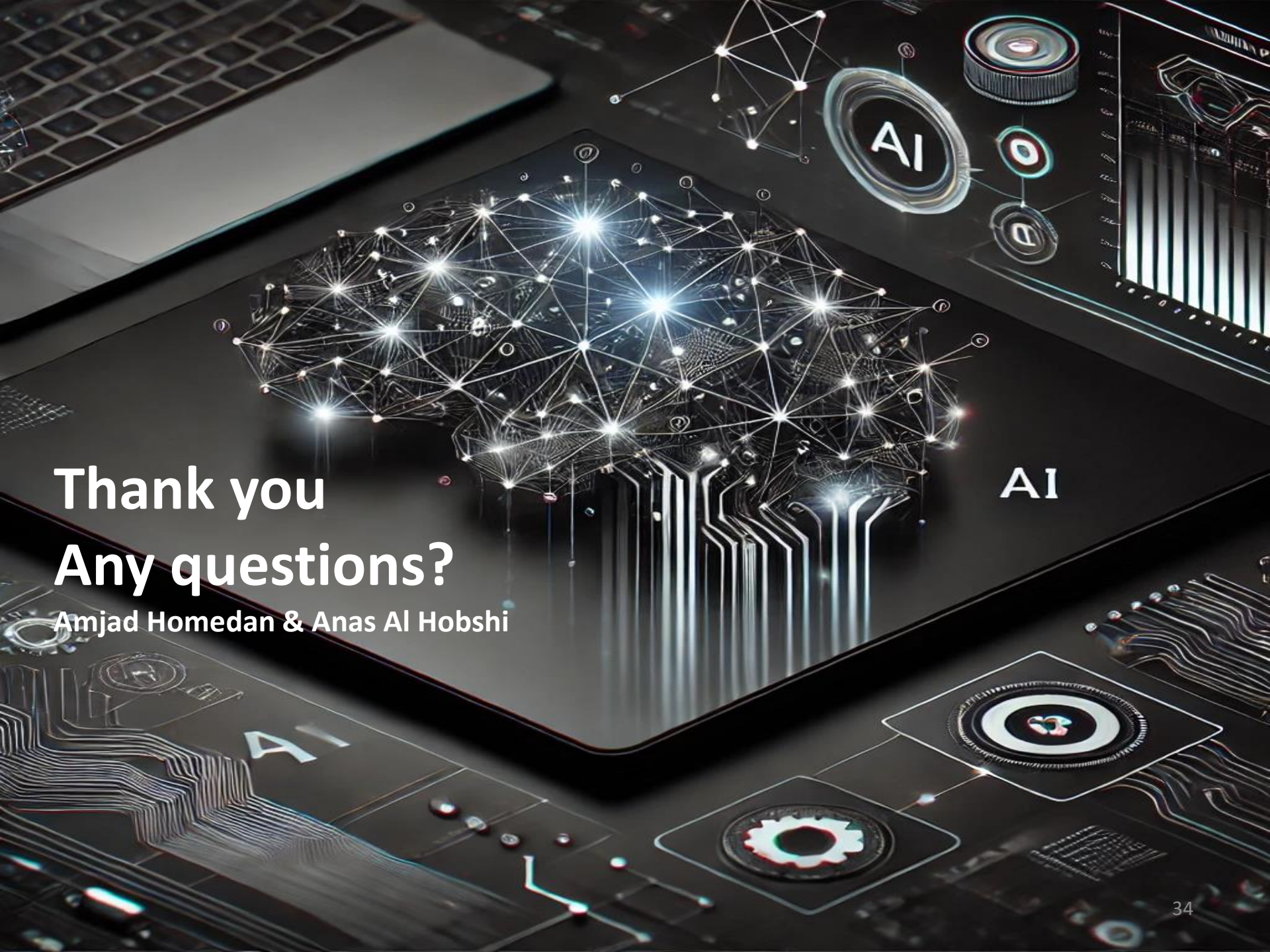
Focuses on enhancing predictive capabilities by incorporating additional features into existing models. The important directions are the integration of project cost prediction using historical cost data, project scope, complexity, and timelines for improving financial forecasting accuracy. Enriching datasets with detailed job descriptions would enhance resource allocation and performance predictions by capturing role-specific contributions to project success.



Adding more project attributes such as pricing, hourly rates or budget categories, would optimize pricing strategies and forecast pricing trends based on project size, duration, and resource needs. This would provide strategic insights into market-aligned pricing and profitability.



Expanding datasets and refining machine learning models with these features would improve cost prediction, resource optimization, and decision-making throughout the project lifecycle, offering comprehensive support for managers and stakeholders.



Thank you
Any questions?

Amjad Homedan & Anas Al Hobshi

AI