





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.

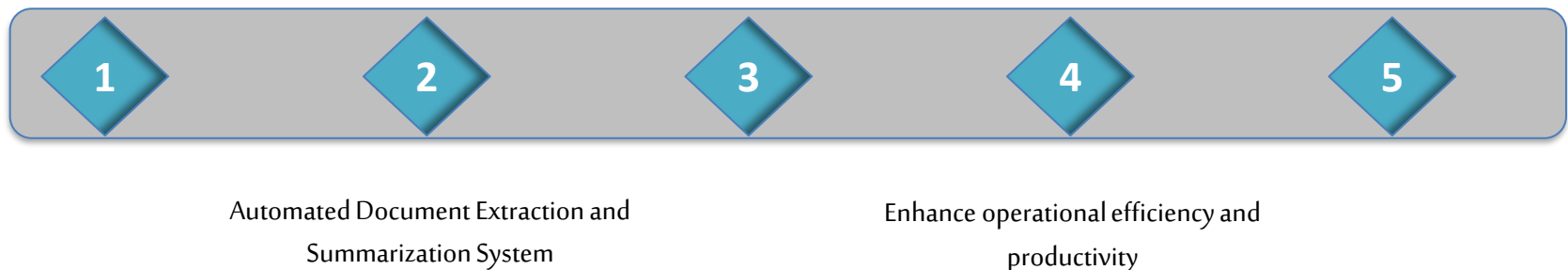


Main Objectives

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.



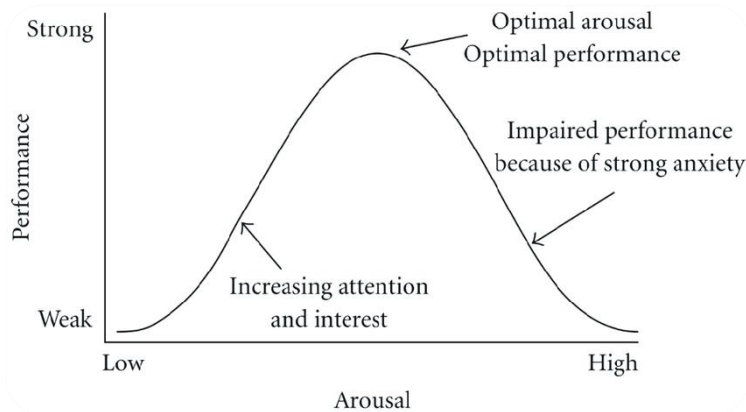
Use Statistical Modeling like MSE, RMSE, R2, Log-Likelihood for evaluation

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

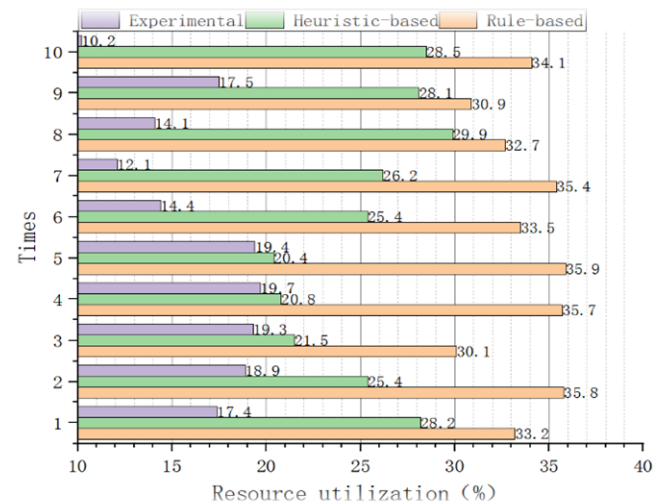
Most of the reference consider the below 3 Principles in their Research

- 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

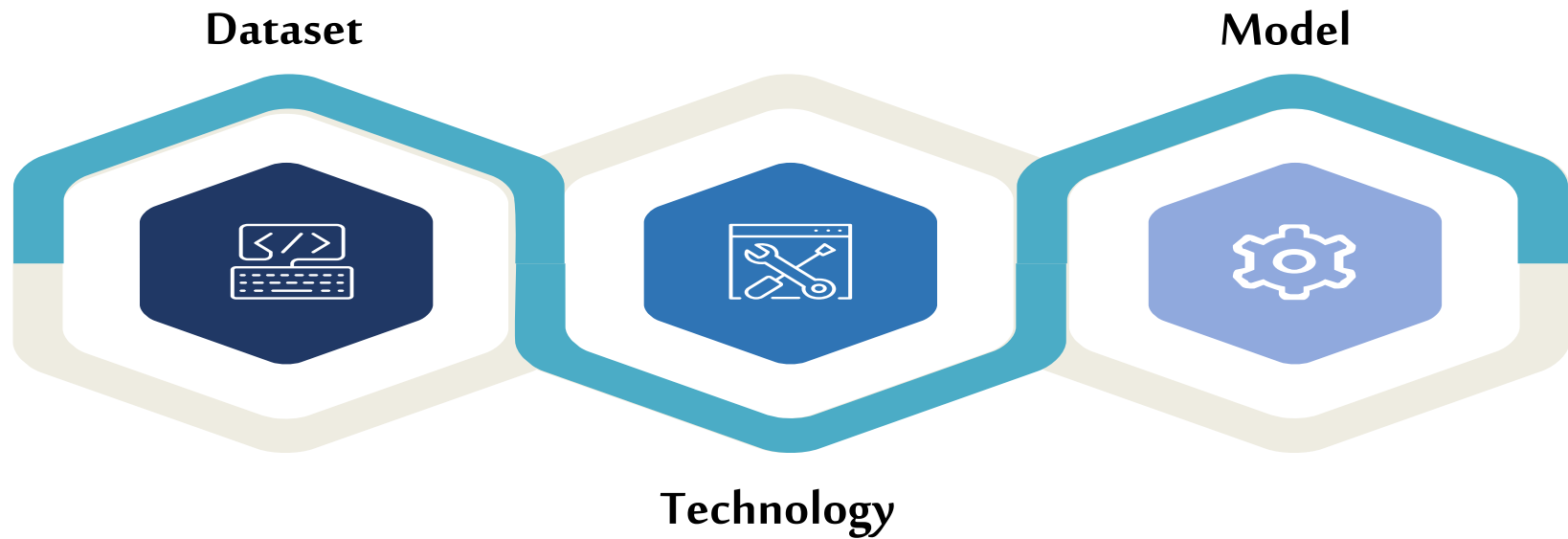
Thesis Methodologies used

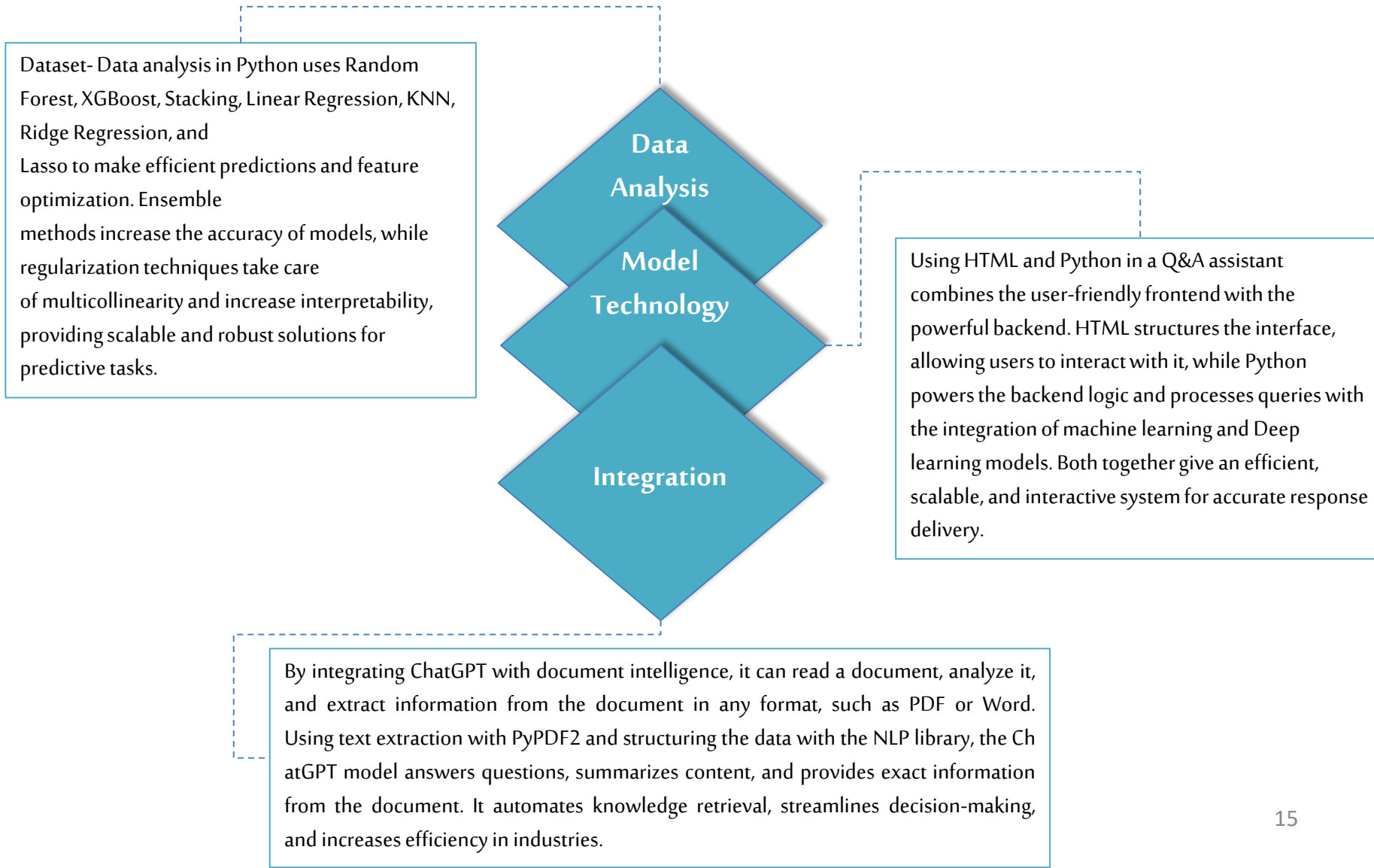
6

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

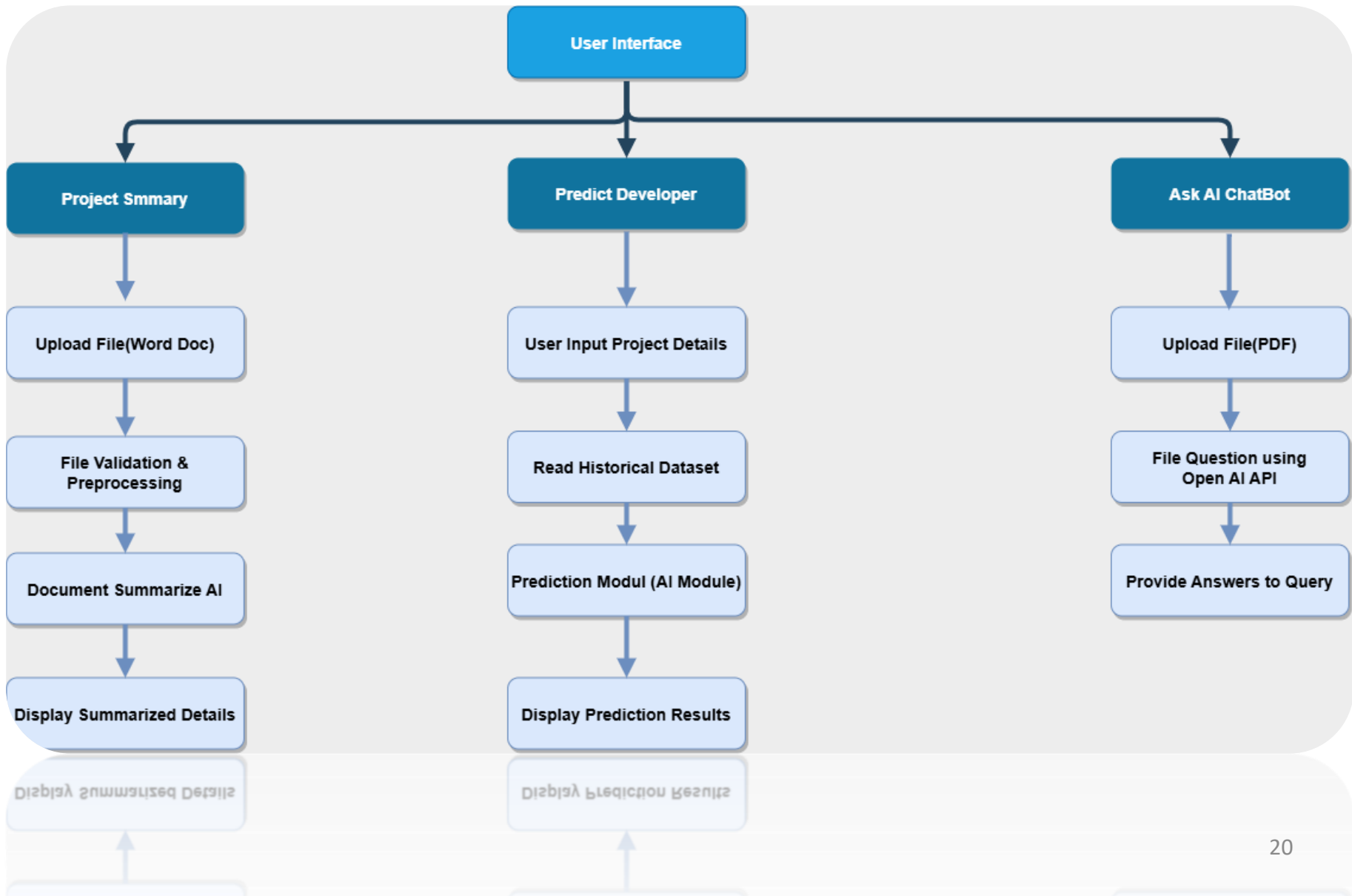
Backend developer count prediction based on project attributes.

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.

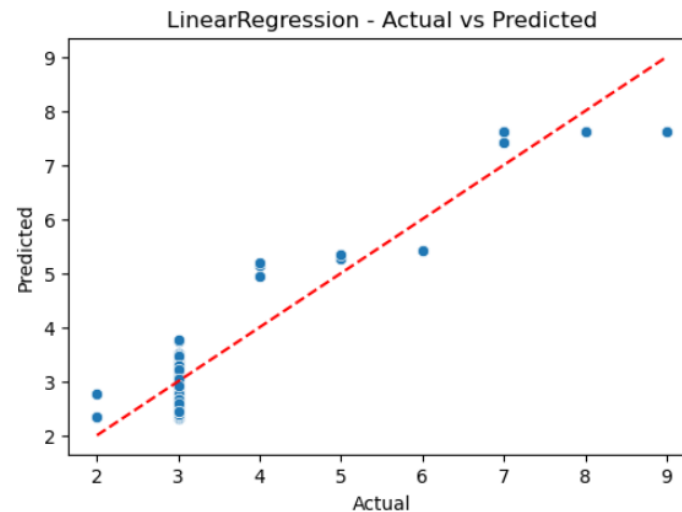
5

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.

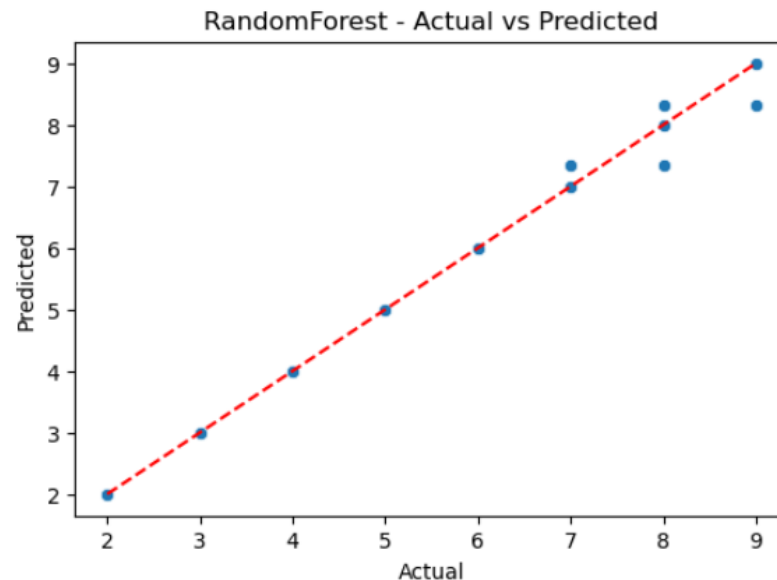
1 Linear Regression

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.



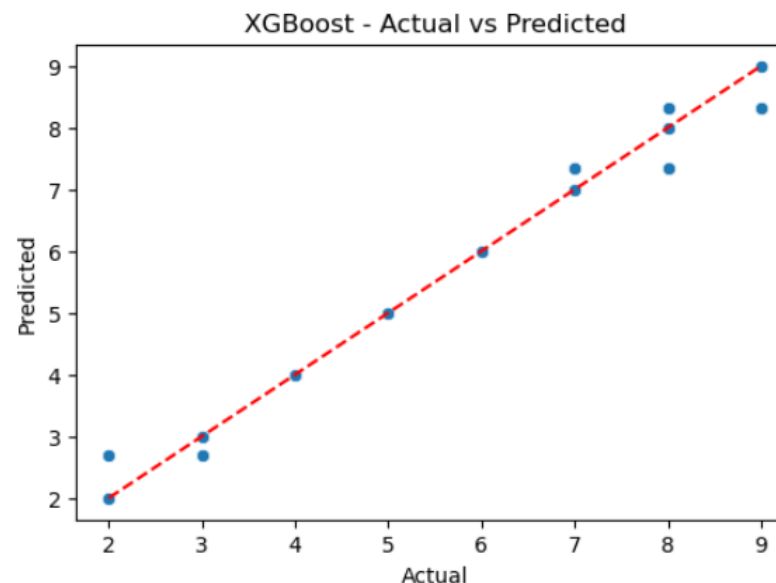
Random Forest

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.



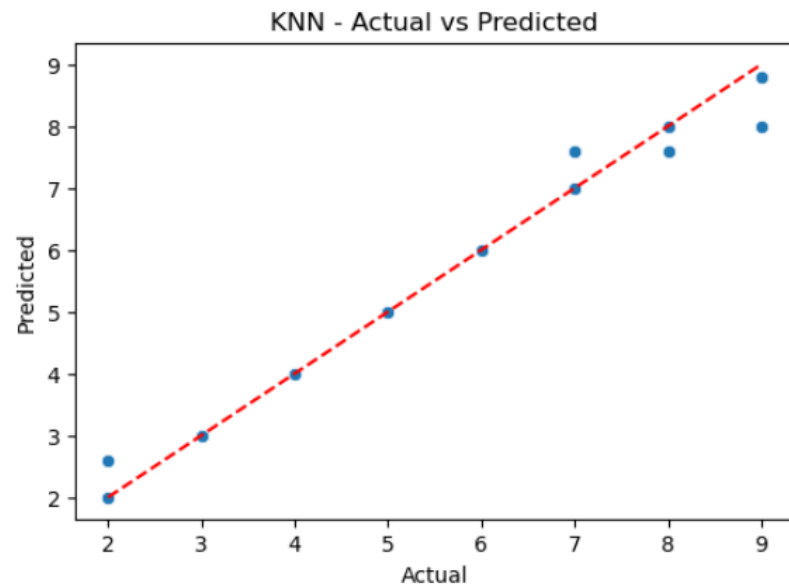
3 XGBoost

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.



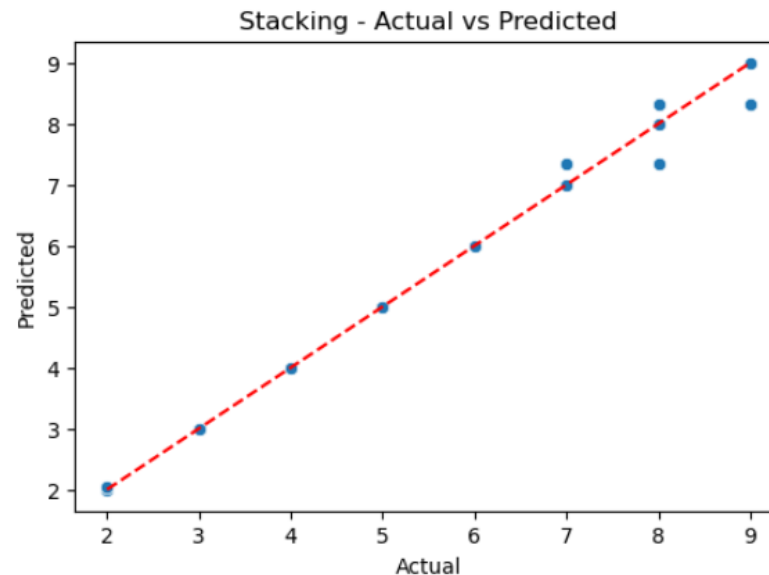
The evaluation metrics for the K-Nearest Neighbors (KNN) model show a Mean Squared Error (MSE) of 0.031152, which is slightly higher compared to XGBoost and Random Forest. The Root Mean Squared Error (RMSE) is 0.176499, indicating a reasonable level of prediction error, though higher than the other models. 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)



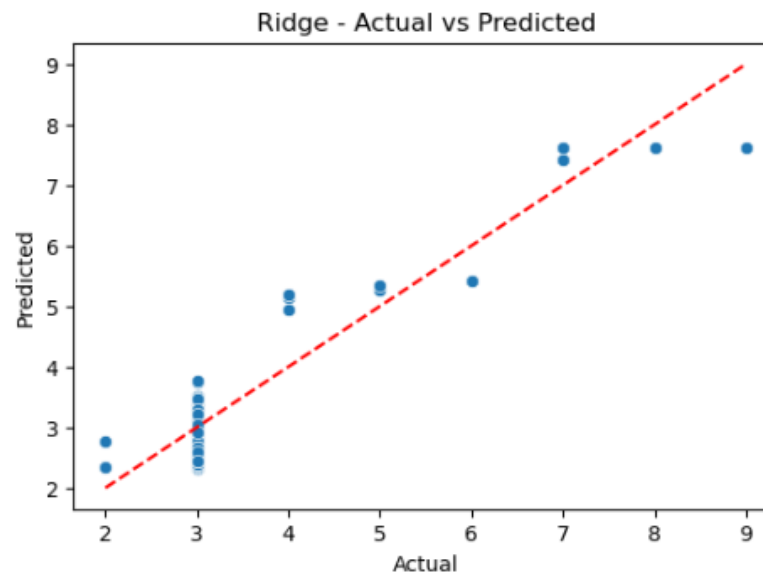
The Stacking model evaluation metrics show an MSE of 0.023846, which is very close to that of the Random Forest model, indicating low prediction error. The RMSE value of 0.154421 also suggests good Acc, with a relatively small deviation between actual and predicted values. The R^2 value of 0.989102 indicates 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



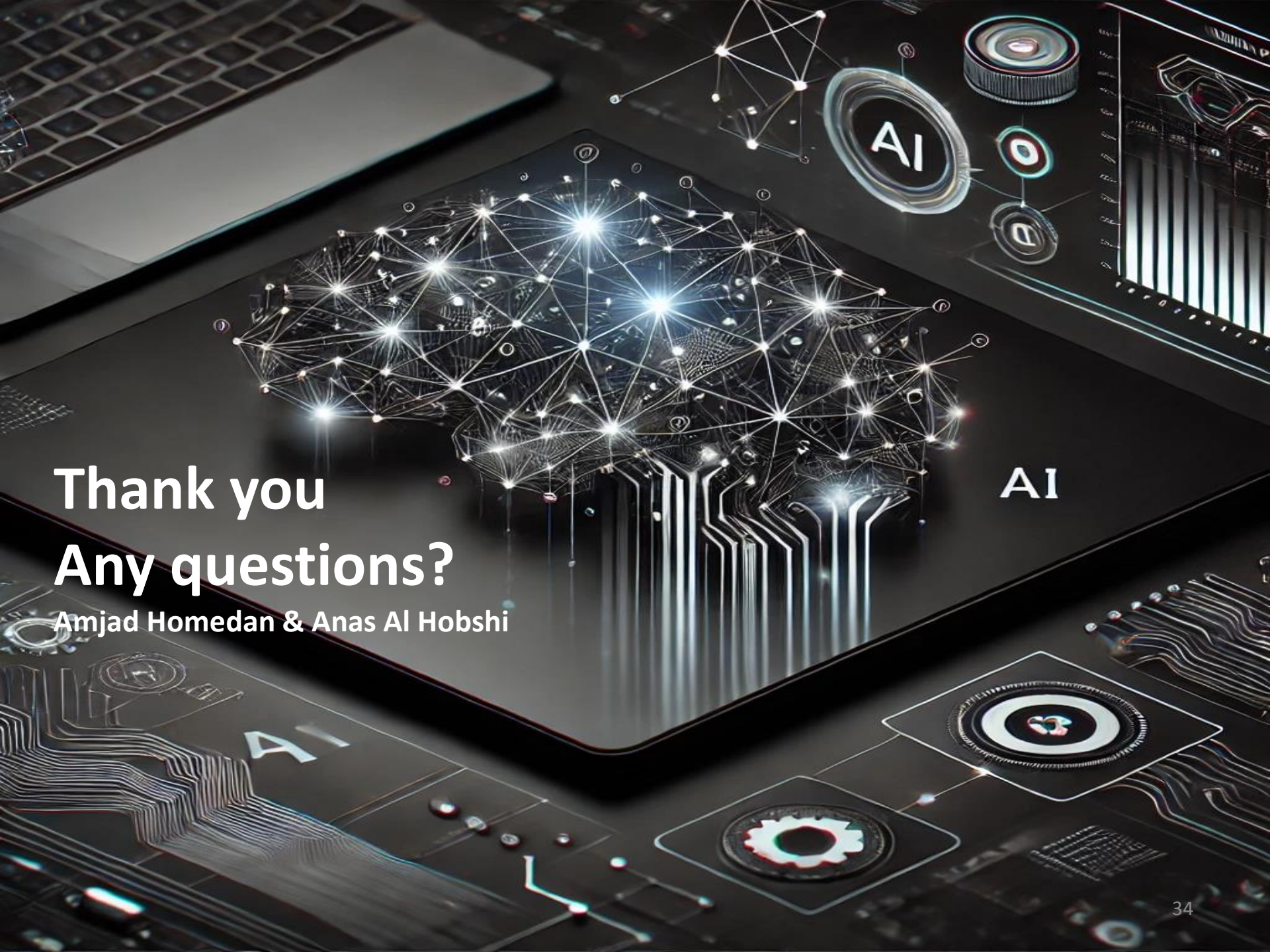
Focuses on enhancing predictive capabilities by incorporating additional features into existing models. The important directions are the integration of project cost prediction using historical cost data, project scope, complexity, and timelines for improving financial forecasting accuracy. Enriching 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



Intelligent Assistant for Resources Optimization and information retrieval

BY

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

Abstract

This thesis explores the integration of machine learning techniques in optimizing resource allocation and enhancing the performance of intelligent assistants. The research proposes a robust framework for real-time resource optimization across various projects by leveraging predictive modeling and data-driven decision-making. By utilizing advanced feature selection methods such as Recursive Feature Elimination (RFE), the study demonstrates how resource management can be optimized through more accurate predictions, thereby enabling smarter and more efficient decision-making. In addition to resource optimization, the thesis investigates the role of intelligent assistants in improving operational outcomes. By integrating machine learning models, including Random Forest, Gradient Boosting, and XGBoost, with Q&A assistants, the study enhances the ability of these systems to process diverse job descriptions and project-related features, leading to better adaptability and accuracy. The research further explores how relationship mining techniques uncover patterns between resource allocation and Q&A performance, offering new insights into improving these systems. The results of the experimental analysis, based on evaluation metrics such as MSE, RMSE, R2, Log-Likelihood, and Cost, indicate significant improvements in both resource optimization and Q&A assistant performance. The proposed models not only enhance the prediction accuracy but also provide a comprehensive approach to the challenges of real-time resource management and QA. This work contributes to the academic field by expanding existing datasets and integrating machine learning into resource optimization processes, offering practical applications in industries aiming to streamline their operations. The thesis outlines future research opportunities, emphasizing the integration of project cost prediction, further expansion of datasets, and refining models for improved adaptability and performance. The findings and proposed framework provide valuable insights for advancing the fields of resource management and intelligent assistant systems.

List of Publications

1. Author(s), Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

2. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

3. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

4. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

5. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

6. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

7. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

8. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

9. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

10. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

11. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

12. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

13. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

14. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

15. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

16. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

17. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

18. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

19. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

20. **Author(s)**, Title of the Paper, *Journal/Conference Name*, Volume (Issue), Year, Pages.

Brief description or impact of the publication, if applicable.

Table of Contents

LIST OF ABBREVIATIONS	IX
LIST OF FIGURES	X
LIST OF TABLES	XI
CHAPTER 1 INTRODUCTION	13
1.1 Problem Definition	13
1.2 Thesis Objectives	14
1.3 Thesis Contribution	15
1.4 Thesis Organization	15
CHAPTER 2 OVERVIEW ON RESOURCES OPTIMIZATION AND INTELLIGENCE QA ASSISTANT	17
2.2 Resources Optimization and Intelligence QA Assistant Data Mining	17
2.4 Goals of Resources Optimization and Intelligence QA Assistant Data Mining	21
2.5 Predicting Resources Optimization and Intelligence QA Assistant Performance	21
CHAPTER 3 RELATED WORK	22
3.1 Introduction	22
3.2 History	22
3.3 Predicting Document Reading and Resources Optimization Using Statistical Analysis	24
3.4 Predicting Document Reading and Resources Optimization Using TensorFlow and Kubernetes	25
3.5 Predicting Document Reading and Resources Optimization Using Machine Learning	27
CHAPTER 4: BACKGROUND	29
4.1 introduction	29
4.2 Conventional Programming vs. Machine Learning	29
4.3 Supervised Learning	29
4.3.1 Regression: Purpose: Predicts continuous values.	30
4.3.2 Classification:	30
4.4 Unsupervised Learning	33
4.5 Semi-supervised Learning	34
CHAPTER 5 PROPOSED APPROACH	35
5.1 Introduction	35
CHAPTER 6 EXPERIMENTAL RESULTS AND DISCUSSION	39
6.1.3.8 Lasso Regression The Lasso model evaluation metrics indicate an MSE of 0.269492, which is very similar to the Ridge model, suggesting a similar level of prediction error. The RMSE of 0.519126 is also comparable to Ridge, indicating a moderate difference between the actual and predicted values. The R^2 value of 0.876838 shows that the Lasso model explains approximately 87.68% of the variance in the target variable, which is close to Ridge but slightly lower. The Log-Likelihood of -2246.482006 is again negative and indicates that the model has residuals that are not perfectly aligned with the assumed distribution, with a worse log-likelihood than Ridge. The Cost of 0.046251 is very similar to Ridge, indicating relatively high inefficiency in minimizing error compared to other models. Overall, the Lasso model has similar performance to Ridge but does not provide significant improvements in terms of accuracy or cost efficiency.	44
6.2 Dealing with Imbalanced Data	45

CHAPTER 7 CONCLUSION AND FUTURE WORK	47
<i>Based on the evaluation of various machine learning models, including Linear Regression, Random Forest, XGBoost, KNN, Stacking, Ridge, and Lasso, it is clear that advanced machine learning (ML) techniques significantly contribute to Resources optimization. The models demonstrated a range of performances in terms of MSE, RMSE, R², Log-Likelihood, and Cost. Random Forest and Stacking emerged as the top-performing models with MSE and RMSE values close to the best possible outcomes, indicating high prediction accuracy. These models minimized errors effectively while also optimizing computational costs.</i>	47
<i>Furthermore, XGBoost showed strong results with a comparable R² value, while KNN and Ridge provided solid performance but with slightly higher error rates, indicating that ensemble models (like Random Forest and Stacking) often provide more robust results in real-world applications.</i>	47
<i>The R² values across models also demonstrated that ML is highly effective at explaining the variance in Resources-related data. With performance nearing or exceeding 87% variance explained for many models, ML has proven to be an efficient tool for improving Resources allocation and decision-making.</i>	47
<i>Machine Learning for Resources Optimization and QA:.....</i>	47
<i>Machine learning has proven to be an invaluable tool for Resources optimization in various fields. By predicting outcomes with high accuracy, ML models enable effective allocation of Resources such as human capital, materials, and time. For example, using models like Random Forest or Stacking, companies can predict demand, optimize schedules, and allocate tasks to the most suitable Resources, leading to improved operational efficiency and cost savings.</i>	47
<i>In the context of Intelligence Q&A Assistant assistants for Resources, ML provides the capability to enhance decision support systems by automating error detection, Resources forecasting, and performance evaluation. With the power of ML models, Intelligence QA Assistants can provide real-time feedback on Resources utilization and performance, ensuring that any discrepancies or inefficiencies are detected early. Additionally, by using predictive modeling, these systems help in anticipating potential issues before they occur, allowing for proactive Resources management and minimized downtime.</i>	47
<i>In conclusion, machine learning not only optimizes Resources allocation but also improves the effectiveness of Intelligence QA Assistants in managing Resources, ensuring smoother operations and better decision-making. By leveraging the power of models like Random Forest, XGBoost, and Stacking, businesses can achieve superior Resources optimization, cost reduction, and overall operational efficiency.....</i>	47
7.3 Future Work.....	48
<i>The future work in this area presents exciting opportunities for further enhancement and expansion of the Resources optimization and predictive capabilities. One key direction is the prediction of project cost, which could be integrated into the existing models to provide more accurate financial forecasting for project management. By incorporating historical cost data and additional features such as the scope of work, complexity, and timelines, the models could predict the total cost of future projects with even higher precision.</i>	48
<i>Moreover, adding more detailed job descriptions to the dataset can help enrich the analysis, enabling the models to better capture the nuances of different roles and responsibilities within projects. This would improve the accuracy of Resources allocation and performance predictions, as more granular job-related data allows for a better understanding of how different job roles contribute to project success and timelines.</i>	48
<i>Another potential area for improvement is the inclusion of project pricing data into the models. By incorporating pricing information, such as hourly rates or project budget categories, the models can optimize pricing strategies, ensuring that project costs are aligned with the value delivered and market conditions. Additionally, predictive models could be used to forecast the pricing trends based on project size, duration, and Resources requirements, providing businesses with strategic insights into competitive pricing strategies and profitability.....</i>	48
<i>By extending the dataset and refining the machine learning models to incorporate these additional features, businesses will be better equipped to predict costs, optimize Resources, and make more informed decisions throughout the project lifecycle. These improvements would not only enhance Resources management but also provide a comprehensive decision-support system for project managers and stakeholders.</i>	48
REFERENCES	49

List of Abbreviations

- **Acc.:** Accuracy
- **ANN:** Artificial Neural Network
- **API:** Application Programming Interface
- **DL:** Deep Learning
- **DM:** Data Mining
- **DT:** Decision Tree
- **KNN:** K-Nearest Neighbor
- **LR:** Logistic Regression
- **ML:** Machine Learning
-

List of Figures

2.1 Educational Data Mining Process	7
2.2 Applications of Educational Data Mining	8
2.3 Distribution of Papers that used different attributes for PSP problem	19
2.4 Student Performance Data Set	21
2.5 Educational Process Mining (EPM): A Learning Analytics Data Set ...	23
3.1 Most Commonly Used ML Algorithms in EDM	35
4.1 Machine Learning Subfields	50
4.2 Supervised Learning	51
4.3 Unsupervised Learning	56
4.4 Reinforcement Learning	59
4.5 Simple ANN	66
4.6 The relation between performances of deep learning algorithm compared to other traditional machine learning algorithms and the amount of data ...	76
5.1 The Proposed Approach	78
6.1 OULAD Database Schema	88
6.2 A Bar Chart Shows Imbalanced Data	91
6.3 Comparing Precisions	107
6.4 Comparing Recalls	108
6.5 Comparing F ₁ -Scores	109
6.6 ROC curve of the ANN model on demographics without using SMOTE	110
6.7 ROC curve of the ANN model on demographics using SMOTE	111
6.8 ROC curve of the ANN model on demographics and scores	111
6.9 ROC curve of the ANN model on demographics, scores, and VLE	112
6.10 ROC curve of the RF model on demographics	112
6.11 ROC curve of the RF model on demographics and scores	113
6.12 ROC curve of the RF model on demographics, scores, and VLE	113
6.13 ROC curve of the k-NN model on demographics	114
6.14 ROC curve of the k-NN model on demographics and scores	114
6.15 ROC curve of the k-NN model on demographics, scores, and VLE	115

6.16 ROC curve of the LR model on demographics	115
6.17 ROC curve of the LR model on demographics and scores	116
6.18 ROC curve of the LR model on demographics, scores, and VLE	116
6.19 ROC curve of the NB model on demographics	117
6.20 ROC curve of the NB model on demographics and scores	117
6.21 ROC curve of the NB model on demographics, scores, and VLE	118
6.22 ROC curve of the SVM model on demographics	118
6.23 ROC curve of the SVM model on demographics and scores	119
6.24 ROC curve of the SVM model on demographics, scores, and VLE	119
6.25 Comparing AUCs	120
6.26 MDI of Demographics	121
6.27 MAD of Demographics	121
6.28 MDI of Demographics and Scores	122
6.29 MAD of Demographics and Scores	123
6.30 MDI of Demographics, Scores, and VLE	124
6.31 MAD of Demographics, Scores, and VLE	124

List of Tables

2.1	Student Performance Data Set Description	21
-----	--	----

2.2	Students' Academic Performance Dataset Description	23
2.3	School_data file fields Description	25
3.1	Summary of researches that used Machine Learning Based models ...	43
6.1	Positions of True Positives, False Positives, True Negatives, and False Negatives in the Confusion Matrix	94
6.2	CM of the ANN Model on Demographics without Using SMOTE ...	96
6.3	CM of the ANN Model on Demographics Using SMOTE	96
6.4	CM of the ANN on Demographics and Scores	97
6.5	CM of the ANN on Demographics, Scores. And VLE	97
6.6	CM of the RF on Demographics only	98
6.7	CM of the RF on Demographics and Scores	98
6.8	CM of the RF on Demographics, Scores and VLE	99
6.9	CM of the k-NN on Demographics	99
6.10	CM of the k-NN on Demographics and Scores	100
6.11	CM of the k-NN on Demographics, Scores, and VLE	100
6.12	CM of the LR on Demographics	101
6.13	CM of the LR on Demographics and Scores	101
6.14	CM of the LR on Demographics, Scores, and VLE	102
6.15	CM of the NB on Demographics	102
6.16	CM of the NB on Demographics and Scores	103
6.17	CM of the NB on Demographics, Scores, and VLE	103
6.18	CM of the SVM on Demographics	104
6.19	CM of the SVM on Demographics and Scores	104
6.20	CM of the SVM on Demographics Scores, and VLE	105
6.21	Comparing Accuracies of Models on Different Data Combinations ...	106
6.22	Important Features Using Filter Methods	125

Chapter 1 Introduction

In today's fast-paced industries, managing resources efficiently and making smart decisions has become crucial. As projects become more complex and data continues to grow, organizations are turning to advanced technologies like machine learning and data mining to make the most of their resources and improve operations. One area that shows promise is combining machine learning-powered resource optimization with intelligent QA (Quality Assurance) systems. These systems have the potential to greatly enhance decision-making processes, allowing businesses to use their resources better and also improve the performance of Q&A assistants. This thesis looks into resource optimization and the role of intelligent QA assistants. Its goal is to create a framework that uses predictive modeling and machine learning techniques to streamline resource allocation across different projects. By integrating advanced methods such as Recursive Feature Elimination (RFE), this research aims to improve the accuracy of resource predictions and Q&A assistant performance. This will lead to smarter decision-making based on data. The findings of this research will benefit industries by improving practices, providing insights into the relationship between resource allocation and QA performance, and offering practical solutions for real-world situations. The main objective of this thesis is to present a comprehensive approach that not only optimizes resource allocation but also enhances the capabilities of intelligent QA assistants using machine learning. By addressing important challenges in resource management and quality assurance, this research seeks to provide a valuable framework for organizations looking to increase productivity, reduce mistakes, and enhance overall operational efficiency. Through this work, we will thoroughly explore the potential of combining machine learning with resource optimization and Q&A assistants. This opens up new possibilities for future advancements in these areas.

1.1 Problem Definition

Resource optimization and intelligent Q&A assistants are indispensable tools for organizations seeking to maximize efficiency, improve decision-making, and enhance customer experience. By leveraging advanced data analytics and artificial intelligence, these systems address critical challenges in resource allocation, task management, and query resolution.

Traditional methods often fall short in effectively managing complex resources, predicting future demands, and providing timely, accurate responses to user queries. Data-driven solutions, on the other hand, offer a more sophisticated approach. By analyzing historical data, identifying patterns, and making data-driven predictions, these systems enable organizations to optimize resource allocation, improve task management, and enhance the performance of Q&A assistants. Key areas where data-driven solutions excel include but not limited to Resource Allocation: Optimizing resource allocation involves matching the right resources to the right tasks, considering factors such as skills, availability, and project priorities. Data-driven systems can analyze historical data to identify optimal resource allocation strategies, predict future resource needs, and proactively address potential bottlenecks.

Task Management: Effective task management requires efficient prioritization, scheduling, and tracking. Data-driven systems can help organizations prioritize tasks based on their importance and urgency, allocate tasks to the most suitable resources, and monitor task progress in real-time. By analyzing historical data, these systems can identify bottlenecks, optimize workflows, and improve overall productivity. Query Resolution: Intelligent Q&A assistants, powered by data-driven insights, can provide

accurate, timely, and relevant responses to user queries. By analyzing historical query data, these systems can identify common patterns, learn from past interactions, and improve their ability to understand and respond to user needs. By harnessing the power of data and advanced analytics, organizations can significantly improve their operational efficiency, reduce costs, and enhance customer satisfaction. As technology continues to evolve, data-driven solutions will play an even more critical role in shaping the future of resource optimization and intelligent Q&A assistants.

1.2 Thesis Objectives

The main goal of this thesis is to find and apply ways to make better use of resources and improve the performance of intelligent Q&A assistants using data-driven models and machine learning methods. The research will primarily focus on creating predictive models that help allocate resources efficiently across different projects and tasks, minimizing waste in the process.

Another important aspect of this study is enhancing the effectiveness of Q&A assistants by employing machine learning and natural language processing techniques. This will result in improved accuracy, scalability, and real-time adaptability of these systems.

To achieve these objectives, we will be using Recursive Feature Elimination (RFE) as a key technique to refine our models. RFE will help us identify the most significant features that contribute to accurate predictions, leading to more efficient resource allocation and better Q&A system performance. Furthermore, we will be exploring the relationship between resource allocation and Q&A system performance through relationship mining techniques. By uncovering hidden patterns in this data, we hope to gain deeper insights into how resources can be managed more effectively for optimal results. Real-time adaptability is another critical focus of this research. We aim to ensure that both our resource optimization models and Q&A assistants have the ability to dynamically adjust to changing conditions in their respective environments. To evaluate the effectiveness and efficiency of our models, we will be using various performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R^2), Log-Likelihood, and Cost. These metrics will provide us with a comprehensive understanding of how well our models are performing. In addition to these existing variables, we also plan on expanding our dataset to include project pricing and job descriptions. This additional information will allow us to conduct a more thorough analysis and improve the accuracy of our model predictions.

Finally, we aim to propose a practical framework for integrating our resource optimization and Q&A assistant models into real-world applications. This framework will enable organizations across various industries to implement these solutions for better resource management and more efficient Q&A processes. Through this research, we hope to make significant contributions towards improving resource optimization strategies and enhancing the capabilities of intelligent Q&A assistants. Our findings may offer valuable insights that can benefit multiple sectors in their pursuit of operational excellence.

1.3 Thesis Contribution

The contributions of this thesis are multifaceted and provide significant advancements in the fields of resource optimization and intelligent Q&A assistants. Firstly, it introduces a robust framework for resource optimization, applying predictive modeling to improve resource allocation across various projects. By incorporating advanced techniques such as Recursive Feature Elimination (RFE), the thesis enhances the accuracy and efficiency of these predictions, enabling smarter, real-time decision-making for resource management. Secondly, it significantly improves Q&A system performance by integrating machine learning models designed to better understand and process a diverse range of job descriptions, project pricing, and related features. This results in Q&A assistants becoming more adaptive, accurate, and scalable, ultimately reducing errors and improving real-time responses.

Additionally, the thesis applies relationship mining techniques to provide valuable insights into the connection between resource allocation and Q&A assistants, uncovering hidden patterns and enabling the development of more intelligent models. The work also contributes to the academic field by expanding datasets with relevant project and job data, which improves model training and enhances prediction accuracy. Furthermore, it advances performance evaluation methodologies by using a comprehensive set of metrics, including MSE, RMSE, R2, Log-Likelihood, and Cost, to ensure a thorough and meaningful assessment of model performance. The practical implications of this research are significant, as the proposed framework for integrating machine learning-driven resource optimization and Q&A systems offers tangible value to industries seeking to streamline operations and boost productivity. Lastly, this thesis lays the foundation for future research, demonstrating how real-time adaptability can be achieved in resource management and intelligent assistant systems. Overall, this work makes a substantial contribution to the fields of data mining, machine learning, and resource management.

1.4 Thesis Organization

Chapter 2 provides a comprehensive overview of resource optimization and intelligent Q&A assistants. It explains the concepts and significance of resource optimization, particularly in industries managing complex projects and diverse human resources. Additionally, it introduces intelligent Q&A systems that utilize machine learning to enhance response accuracy and efficiency. The chapter explores how integrating these two domains can significantly improve productivity and enable smarter decision-making. By leveraging data-driven methods, organizations can optimize resources in real-time, enhance Q&A assistant performance, and improve overall operational outcomes.

Chapter 3 delves into the existing body of work related to resource optimization, intelligent QA assistants, and their real-world applications. It reviews literature on predictive modeling techniques, machine learning models for resource allocation, and the use of artificial intelligence in Q&A assistants. The chapter also highlights critical methodologies, approaches, and challenges in these fields. A key focus is on identifying gaps that this thesis seeks to address, particularly in integrating machine learning-driven resource optimization with intelligent Q&A assistants to improve both efficiency and accuracy.

Chapter 4 serves as the theoretical foundation for understanding the research presented in this thesis. It covers essential topics such as machine learning fundamentals, supervised learning principles, feature selection techniques like Recursive Feature Elimination (RFE), and the use of various algorithms (e.g., Random Forest, Gradient Boosting, XGBoost). It also explains the importance of performance evaluation metrics such as MSE, RMSE, R2, Log-Likelihood, and Cost, helping readers understand the methodologies and theories applied throughout the research and setting the stage for the proposed solutions.

Chapter 5 outlines the novel framework introduced in this thesis for resource optimization and intelligent QA assistants. It details the approach of combining predictive modeling, data mining, and machine learning to optimize resource allocation across various projects. The chapter explains how Recursive Feature Elimination (RFE) is used to select the most relevant features for model training, enhancing prediction accuracy. It also highlights the integration of multiple machine learning models, including Linear Regression, Random Forest, and Stacking Regressors, to tackle challenges in both resource optimization and Q&A assistant performance.

Chapter 6 presents the experimental results evaluating the performance of the proposed models. It compares the results of various machine learning models, such as Linear Regression, Gradient Boosting, Random Forest, XGBoost, KNN, Ridge, and Lasso, using performance metrics like MSE, RMSE, R2, Log-Likelihood, and Cost. The chapter provides an in-depth discussion on how these models perform in optimizing resource allocation and improving QA assistant efficiency. The findings are discussed in terms of their contribution to enhancing real-time decision-making and operational outcomes.

Chapter 7 concludes the thesis by summarizing the key findings and contributions to resource optimization and intelligent Q&A assistants. It discusses how the proposed framework can be implemented in real-world scenarios to enhance efficiency, accuracy, and scalability. The chapter also outlines potential directions for future research, including the integration of project cost prediction and the expansion of the dataset to include more job descriptions. It suggests further refinements of the models to improve performance and adaptability, paving the way for continued advancements in this area.

Chapter 2 Overview on Resources Optimization and Intelligence QA Assistant

2.1 Introduction

In today's fast-paced and competitive environment, organizations face increasing pressure to maximize efficiency and minimize waste. Resource optimization has emerged as a critical solution to address these challenges by enabling the effective allocation and utilization of resources, whether human, financial, or technical. By leveraging advanced technologies like machine learning (ML) and artificial intelligence (AI), businesses can make data-driven decisions that improve productivity and reduce costs. An integral part of this innovation is the development of an Intelligent QA Assistant, which serves as a proactive tool to streamline workflows and enhance decision-making processes. These assistants are equipped with AI-driven capabilities to automate repetitive tasks, provide accurate recommendations, and ensure quality control. By integrating resource optimization with intelligent Q&A assistants, organizations can not only enhance operational efficiency but also improve project outcomes through predictive analysis and tailored recommendations. This overview highlights the transformative role of resource optimization and intelligent QA assistants in addressing modern business challenges, emphasizing the synergy between cutting-edge technology and efficient resource management. Together, they pave the way for smarter, more agile, and sustainable operations.

2.2 Resources Optimization and Intelligence QA Assistant Data Mining

The fusion of resource optimization and intelligent QA assistants is revolutionizing how organizations leverage data mining to enhance operational efficiency. Data mining, the process of extracting valuable insights from vast datasets, serves as the backbone of intelligent systems designed to allocate resources effectively and automate quality assurance processes. Resource optimization focuses on identifying and utilizing the most efficient methods to allocate human, financial, and material resources. Data mining techniques such as clustering, regression, and predictive modeling allow businesses to analyze historical and real-time data, uncover patterns, and predict future resource needs. This ensures that resources are allocated where they are most impactful, reducing waste and improving overall productivity. On the other hand, intelligent QA assistants harness data mining to automate repetitive quality control tasks and provide actionable insights. These systems employ machine learning algorithms to identify potential risks, detect anomalies, and recommend solutions, ensuring that projects meet the highest standards. By analyzing project data, such as resource utilization rates, task completion times, and error patterns, these assistants enable businesses to proactively address issues and maintain quality benchmarks. The integration of data mining within resource optimization and Q&A assistants not only improves decision-making but also empowers organizations to adapt to dynamic challenges. This synergy unlocks new possibilities for efficiency, scalability, and precision, marking a significant step forward in the evolution of intelligent enterprise solutions.

2.2.1 Prediction

Prediction is at the core of enhancing resource optimization and intelligent QA assistant systems. By leveraging advanced predictive analytics, organizations can forecast future trends, anticipate challenges, and make proactive decisions, leading to improved efficiency and accuracy.

In resource optimization, prediction models analyze historical data to forecast project demands, resource utilization, and workforce requirements. For example, machine learning algorithms such as regression and time-series forecasting can predict project costs, timelines, and resource shortages, enabling businesses to allocate resources more effectively. These predictions help organizations prevent bottlenecks, reduce downtime, and ensure the optimal deployment of human and material assets.

For intelligent QA assistants, prediction enhances quality assurance by identifying potential risks and errors before they occur. Predictive algorithms can analyze patterns in defect rates, system performance, and compliance metrics to flag anomalies or trends that may compromise quality. This allows Q&A assistants to recommend corrective actions early, reducing rework costs and maintaining high-quality deliverables.

By integrating predictive capabilities, resource optimization and Q&A assistants transition from reactive to proactive. This shift empowers organizations to anticipate future needs, streamline operations, and achieve superior outcomes, driving innovation and competitiveness in a data-driven world.

2.2.2 Clustering

Clustering is a powerful unsupervised machine learning technique used to group data points with similar characteristics into clusters. In the context of resource optimization and intelligent QA assistants, clustering provides valuable insights by identifying hidden patterns and enabling data-driven decision-making.

In resource optimization, clustering helps categorize projects, teams, or tasks based on their similarities. For example, projects can be grouped by complexity, budget, or required skill sets. This grouping aids in better resource allocation by matching teams with appropriate tasks, reducing inefficiencies, and improving project outcomes. Additionally, clustering can identify underutilized resources or areas with excessive demand, allowing organizations to balance workloads effectively.

For intelligent QA assistants, clustering is used to detect anomalies, segment defects, and group similar queries or issues. By clustering common defects, Q&A assistants can prioritize addressing recurring problems, leading to faster resolution and improved quality. Similarly, customer support data can be clustered to identify frequently asked questions, enabling QA assistants to provide quick and accurate responses.

Clustering enhances the ability to discover patterns that may not be immediately apparent, enabling smarter resource management and quality assurance processes. It facilitates a deeper understanding of data relationships, supports proactive strategies, and optimizes operations, driving higher efficiency and effectiveness in organizations.

2.2.3 Relationship Mining

Relationship mining, a data mining technique, uncovers hidden connections, correlations, and patterns within datasets. This powerful tool significantly impacts resource optimization and intelligent QA assistants. In resource optimization, relationship mining identifies task dependencies, optimizes resource utilization, and enables predictive analytics, leading to informed decision-making and efficient resource allocation. For instance, analyzing historical data can reveal optimal team configurations and predict future resource needs. In intelligent QA assistants, relationship mining enhances user experiences by analyzing customer queries, identifying defect root causes, and predicting user behavior. This enables proactive problem-solving and personalized support. By unlocking the hidden potential within data, relationship mining drives innovation, efficiency, and strategic decision-making in both resource management and quality assurance domains.

2.3. Discovery with Model

Model-driven discovery is a powerful data-driven approach that leverages machine learning and statistical methods to uncover valuable insights from complex datasets. This technique has significant applications in domains such as resource optimization and intelligent QA assistants, where it can significantly improve efficiency, decision-making, and user experience.

In resource optimization, model-driven discovery enables organizations to make data-driven decisions by identifying patterns and trends in historical data. This includes predicting future resource demand, optimizing workforce allocation, and forecasting project costs. For instance, by analyzing past project data, organizations can identify optimal resource configurations and predict potential bottlenecks.

In the realm of intelligent QA assistants, model-driven discovery empowers systems to understand user intent, detect errors, and personalize responses. By analyzing user interactions and historical data, QA assistants can provide more accurate and relevant information. For example, a QA assistant can use machine learning to classify user queries and identify the most appropriate response. A variety of machine learning models are employed in model-driven discovery, including regression models, clustering models, decision trees, random forests, and neural networks. Each model is suited to different tasks, such as predicting numerical values, grouping similar data points, classifying data into categories, and recognizing complex patterns.

By implementing model-driven discovery, organizations can reap numerous benefits, including increased efficiency, improved decision-making, and enhanced user experiences. By harnessing the power of data and machine learning, organizations can unlock new opportunities and drive innovation.

2.4 Distillation of Data for Human Judgment

In our data-driven world, the ability to extract meaningful insights from vast amounts of information is paramount. Data distillation, the process of transforming raw data into concise, actionable insights, plays a crucial role in empowering human decision-making. By refining complex datasets, data distillation enables individuals to make informed judgments without being overwhelmed by information overload. Key processes involved in data distillation include data aggregation, feature selection, data summarization, and pattern detection. Data aggregation involves collecting data from various sources and consolidating it into a unified dataset. Feature selection identifies the most relevant attributes, discarding redundant or irrelevant information. Data summarization transforms complex datasets into concise summaries using statistical measures, visualizations, and clustering techniques. Finally, pattern detection utilizes machine learning algorithms to uncover trends, anomalies, and correlations within the data. Data distillation has numerous applications in various domains, particularly in resource optimization and intelligent Q&A assistants. By distilling complex datasets, organizations can make informed decisions about resource allocation, identify potential bottlenecks, and optimize workflows. Intelligent Q&A assistants can leverage data distillation to provide more accurate and timely responses to user queries, enhancing customer satisfaction and operational efficiency.

The benefits of data distillation are manifold. It simplifies complex datasets, reduces the time required to interpret data, and focuses attention on the most critical factors. By providing clear and concise insights, data distillation empowers individuals to make faster, more accurate, and better-informed decisions. In conclusion, data distillation is not about reducing information but rather about enhancing its usability and interpretability. By bridging the gap between raw data and actionable insights, data distillation empowers individuals to make informed decisions, improve efficiency, and drive innovation. As the volume and complexity of data continue to grow, data distillation will become an increasingly essential tool for navigating the information age.

2.4 Goals of Resources Optimization and Intelligence QA Assistant Data Mining

Data mining has emerged as a powerful tool for driving efficiency and innovation in resource optimization and intelligent Q&A assistants. By delving into vast datasets, data mining techniques uncover hidden patterns, correlations, and actionable insights that empower organizations to make informed decisions and optimize their operations. In the realm of resource optimization, data mining enables organizations to analyze historical data, identify trends, and forecast future demand. By understanding the factors that influence resource utilization, organizations can optimize staffing levels, allocate resources effectively, and minimize costs. For instance, by analyzing past project data, organizations can identify the optimal resource allocation for different project types, ensuring efficient utilization of resources. In the domain of intelligent Q&A assistants, data mining enhances the capabilities of AI-powered assistants by enabling them to learn from past interactions and provide more accurate and personalized responses. By analyzing customer support tickets, chat logs, and other relevant data, these systems can identify frequently asked questions, automate responses, and prioritize critical issues. Additionally, data mining can help identify underlying causes of common problems, enabling organizations to implement preventive measures and improve overall customer satisfaction. The integration of data mining into resource optimization and intelligent Q&A assistants fosters a data-driven culture, driving innovation and improving operational efficiency. By harnessing the power of data, organizations can gain a competitive edge, enhance customer satisfaction, and achieve sustainable growth.

2.5 Predicting Resources Optimization and Intelligence QA Assistant Performance

Predicting the performance of resource optimization and intelligent Q&A assistants is paramount for maximizing efficiency and ensuring reliable decision-making. By leveraging predictive analytics and machine learning models, organizations can proactively anticipate system behaviors, identify potential bottlenecks, and refine operational strategies. Key areas of focus for predicting performance include, Resource Utilization Forecasting: Accurately forecasting future resource demand is crucial for optimizing resource allocation. By analyzing historical data, identifying trends, and considering external factors such as market fluctuations or seasonal variations, organizations can predict peak demand periods, potential shortages, and surplus resources. This enables proactive resource planning, preventing bottlenecks and ensuring optimal resource utilization.

Performance Metric Tracking and Benchmarking: Monitoring key performance indicators (KPIs) such as response time, accuracy, and efficiency is essential for evaluating system performance. By establishing relevant benchmarks and tracking KPIs over time, organizations can identify areas for improvement, measure the impact of changes, and ensure alignment with organizational goals. Advanced analytics techniques, such as time series analysis and statistical process control, can help identify trends, anomalies, and potential issues.

Query Resolution Success Prediction: Predicting the probability of successful query resolution is essential for managing user expectations and optimizing resource allocation. By analyzing historical query data, user behavior patterns, and system performance metrics, organizations can identify complex queries, estimate resolution time, and prioritize critical issues. Machine

learning models, such as decision trees and random forests, can help classify queries based on their complexity and assign them to the appropriate resources. **Project Duration and Cost Prediction:** Accurate project forecasting is crucial for effective planning, budgeting, and resource allocation. By analyzing historical project data, considering factors such as project complexity, resource availability, and external factors, organizations can estimate project timelines and costs with greater accuracy. Machine learning models, such as regression analysis and neural networks, can help identify the most influential factors and predict project outcomes with higher precision.

Identifying Improvement Opportunities: Continuous improvement is essential for optimizing system performance. By analyzing historical data, identifying outliers, and using statistical techniques, organizations can identify areas where system performance can be enhanced. This may involve optimizing resource allocation, refining query resolution processes, or improving system configuration. **Optimizing Feedback Loops:** Integrating feedback into the prediction process is crucial for refining model accuracy and ensuring alignment with evolving organizational needs. By continuously monitoring system performance, collecting user feedback, and retraining models with updated data, organizations can improve the accuracy and reliability of their predictions.

By effectively addressing these key areas, organizations can proactively manage resource optimization and intelligent Q&A assistants, leading to improved decision-making, enhanced efficiency, and increased customer satisfaction.

Chapter 3 Related Work

3.1 Introduction

3.2 History

The integration of artificial intelligence (AI) into human Resources management (HRM) represents a major evolution in the field, reflecting broader historical shifts in economic organization and technological advancements. Traditionally, HRM relied on manual processes for recruitment, performance management, and employee relations, but the advent of AI has introduced automation and data-driven decision-making, enhanced efficiency and reducing biases in recruitment (Jia et al., 2018; Vivek & Yawalkar, 2019). AI technologies like machine learning and data mining now enable organizations to analyze large amounts of data for strategic planning and performance evaluation, fundamentally transforming HR functions (Jia et al., 2018). This transition not only streamlines HR activities but also requires HR professionals to develop new skills to effectively interact with AI systems (Jia et al., 2018); (Sangeeta, 2019a). Similarly, the history of question answering (QA) in AI, which began with early systems like ELIZA, has evolved from simple rule-based approaches to advanced models leveraging machine learning and neural networks, such as BERT and DistilBERT (Pearce et al., 2023). These advancements, including the integration of external knowledge bases and representation learning, have improved the accuracy of responses and reflect the growing complexity of processing natural language and generating human-like answers (Goel et al., 2022a). Both HRM and QA illustrate how AI's historical and technological progress is shaping the future of workforce management and natural language processing.

3.3 Predicting Document Reading and Resources Optimization Using Statistical Analysis

Resources optimization in organizations can be significantly enhanced through various statistical analysis methods, particularly in the context of human Resources management. For instance, the use of process mining techniques allows for the analysis of event logs to understand Resources behavior and improve allocation strategies, as demonstrated in studies that apply statistical models to optimize Resources distribution based on performance metrics and constraints (Yeon et al., 2022). Additionally, the integration of data analytics in business processes supports the identification of patterns in employee performance, which can inform better decision-making regarding Resources allocation (Votto et al., 2021a). By leveraging accurate data, organizations can align their Resources more effectively with operational needs, ultimately driving productivity and efficiency (Sangeeta, 2019a). Thus, statistical analysis serves as a critical tool in the ongoing effort to optimize Resources within organizations. In addition, some authors highlighted the relationship between the workload and Performance of the workers. (Yeon et al., 2022). Figure 3.1, and

Statistical analysis plays a crucial role in the development and evaluation of question answering (QA) systems. Various QA methodologies utilize statistical measures such as exact match (EM), F1 score, in context also Author referred to the relationship between process performance and Resources allocation.

(Yeon et al., 2022)

mean reciprocal rank (MRR) to assess the accuracy and effectiveness of their responses, particularly in datasets like SQuAD (Nuamah, 2021a). Additionally, modern Q&A assistants have evolved from simple rule-based approaches to more sophisticated models that incorporate statistical reasoning and machine learning techniques, allowing for improved performance in retrieving and generating answers (Nuamah, 2021a). The integration of statistical methods enhances the interpretability and efficiency of Q&A assistants, enabling them to handle complex queries across diverse modalities, including text and multimedia (Nuamah, 2021a). Overall, statistical analysis is integral to refining Q&A assistants and ensuring their reliability in providing accurate information.

Resources optimization in organizations can be significantly enhanced through various statistical analysis methods, particularly in the context of human Resources management. For instance, the use of process mining techniques allows for the analysis of event logs to understand Resources behavior and improve allocation strategies, as demonstrated in studies that apply statistical models to optimize Resources distribution based on performance metrics and constraints (Yeon et al., 2022). Additionally, the integration of data analytics in business processes supports the identification of patterns in employee performance, which can inform better decision-making regarding Resources allocation (Votto et al., 2021a). By leveraging accurate data, organizations can align their Resources more effectively with operational needs, ultimately driving productivity and efficiency (Sangeeta, 2019a). Thus, statistical analysis serves as a critical tool in the ongoing effort to optimize Resources within organizations. In addition, some authors highlighted the relationship between the workload and Performance of the workers. (Yeon et al., 2022). Figure 3.1, and

Statistical analysis plays a crucial role in the development and evaluation of question answering (QA) systems. Various QA methodologies utilize statistical measures such as exact match (EM), F1 score, in context also Author referred to the relationship between process performance and Resources allocation.

(Yeon et al., 2022)

mean reciprocal rank (MRR) to assess the accuracy and effectiveness of their responses, particularly in datasets like SQuAD(Nuamah, 2021a). Additionally, modern Q&A assistants have evolved from simple rule-based approaches to more sophisticated models that incorporate statistical reasoning and machine learning techniques, allowing for improved performance in retrieving and generating answers (Nuamah, 2021a). The integration of statistical methods enhances the interpretability and efficiency of Q&A assistants, enabling them to handle complex queries across diverse modalities, including text and multimedia(Nuamah, 2021a). Overall, statistical analysis is integral to refining Q&A assistants and ensuring their reliability in providing accurate information.

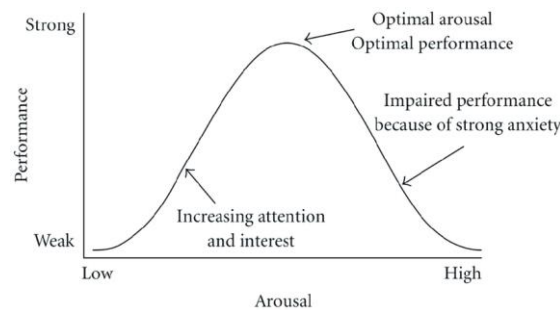


Figure 3.1 Hebbian version of the Yerkes-Dodson law

3.4 Predicting Document Reading and Resources Optimization Using TensorFlow and Kubernetes

This study presents significant contributions, including an artificial intelligence-based Resources scheduling method that utilizes TensorFlow and CooperNet, tested and validated through experiments on an open controller. By establishing a Resources scheduling model and optimization algorithm, the study achieves intelligent Resources allocation and scheduling, evidenced by experimental results showing an improved scheduling efficiency of 10.2% and a response time reduced to 202 milliseconds—a marked improvement over traditional methods. As AI technology continues to evolve, more applications increasingly rely on AI-powered systems. For instance, in OpenEuler, TensorFlow configuration requires several key parameters, including model selection, input data from OpenStackDataFrame (which may be segmented for large datasets), and additional parameters to optimize training and testing, such as training set sizes and activation functions. Additionally, a TensorFlow model repository is maintained for efficient model management and monitoring, deployed through Kubernetes. Resources scheduling on Kubernetes is managed by implementing YARN (Yet Another Resources Negotiator) standards, where a YARN node is established with `yarn.fit_layer_selection_data` selected as the Resources scheduler. This scheduler does not run as a traditional Kubernetes container; instead, it handles Resources scheduling through Kubernetes functions. The article also delves into

Resources scheduler optimization formulas, primarily focusing on hardware Resources management, which may also find applications in HRD systems. With OpenEuler built on OpenStack, this approach ensures seamless accessibility to required computing Resources, facilitating reliable Resources allocation for applications running on this platform.(Li & Zhang, 2024).

In the first experiment, as below Figure 3.2, the Resources utilization rate of the rule-based Resources scheduling method was 33.2%, peaking at 35.9% in the fifth experiment and recording 34.1% in the tenth experiment, with the lowest value of 30.1% observed in the third experiment. For the heuristic-based Resources scheduling method, the utilization rate started at 28.2% in the first experiment, dropped to a minimum of 20.4% in the fifth, and reached 28.5% in the tenth, with the highest rate of 29.9% occurring in the eighth experiment. The Resources scheduling method introduced in this study showed a utilization rate of 17.4% in the first experiment, increased to 19.4% in the fifth, reached a minimum of 10.2% in the tenth, and a maximum of 19.7% in the fourth experiment. These results indicate that the proposed Resources scheduling method achieves a lower Resources utilization rate, effectively minimizing system burden while maintaining smooth task execution.(Li & Zhang, 2024).

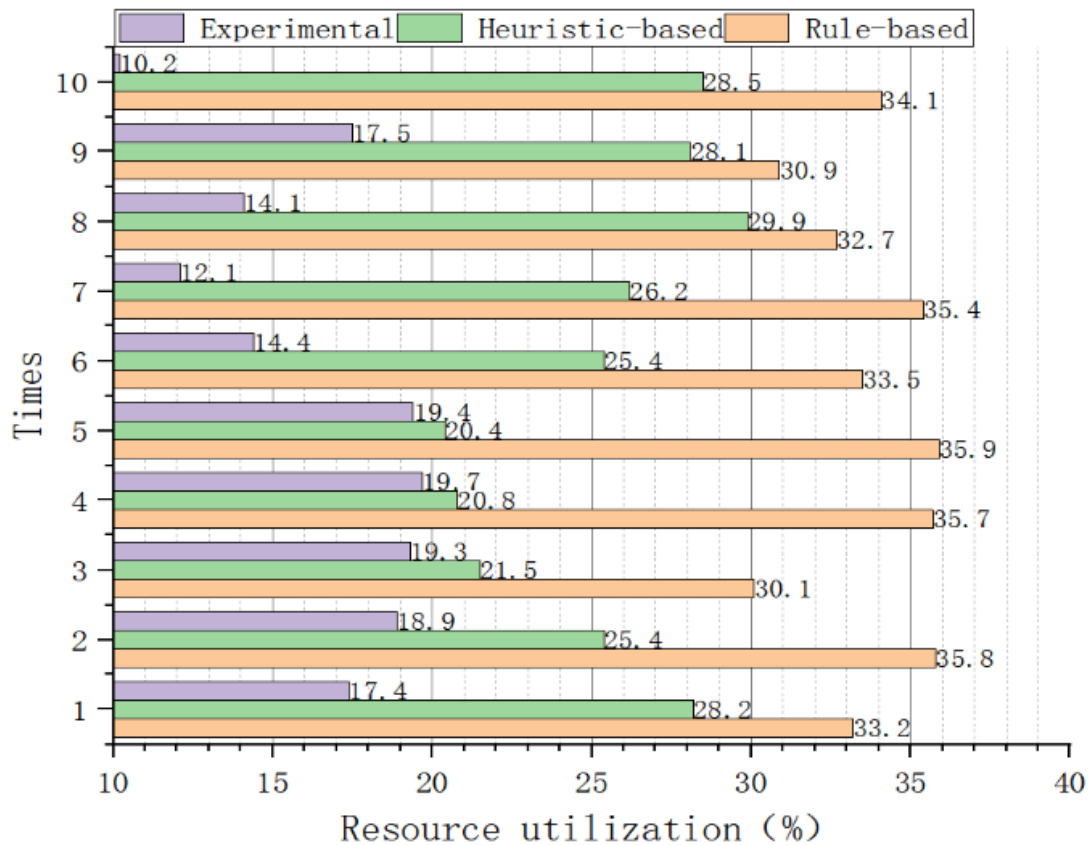


Figure 3.2 Resources Utilization Rate

3.5 Predicting Document Reading and Resources Optimization Using Machine Learning

Resources optimization in organizations can be significantly enhanced through the application of machine learning techniques, particularly in the context of human Resources allocation. Various studies highlight the use of process mining and reinforcement learning to analyze operational data and improve decision-making regarding Resources distribution. For instance, one approach involves mining event logs to uncover patterns that inform Resources allocation strategies, thereby optimizing workflow efficiency and reducing execution time under specific constraints (Yeon et al., 2022). Additionally, the integration of human-centric networks allows for a more nuanced understanding of employee roles and their centrality within processes, which can further refine Resources allocation decisions (Yeon et al., 2022). By leveraging these advanced methodologies, organizations can streamline their human Resources management processes and enhance overall operational effectiveness (Votto et al., 2021a; Yeon et al., 2022) .

Machine learning plays a crucial role in enhancing question and answering (QA) systems by enabling them to process natural language queries and extract relevant information effectively. For instance, a hybrid classification system, as described in one study, utilizes both machine learning and semantic rules to interpret user questions and generate structured queries for knowledge bases, thereby improving response accuracy (Goel et al., 2022a). Additionally, the integration of deep learning techniques is emphasized for their ability to automate reasoning processes in QA, although they are often complemented by symbolic methods to achieve better interpretability and robustness (Nuamah, 2021a). Furthermore, AI systems can automate repetitive tasks in QA, such as candidate screening in recruitment, showcasing the practical applications of machine learning in real-world scenarios (Sangeeta, 2019b). Overall, the combination of machine learning with traditional methods enhances the efficiency and effectiveness of Q&A assistants across various domains.

3.6 Predicting Document Reading and Resources Optimization Using Deep Learning:

The integration of deep learning techniques into Resources optimization is highlighted through various approaches, particularly in human Resources management. One method involves using reinforcement learning to develop policies that optimize Resources allocation during business process execution, as discussed in the context of human-centric networks. This approach emphasizes the importance of analyzing operational event logs to understand the relationships and centrality metrics of employees, which can inform more effective Resources distribution. Additionally, the evolution of information technology and business process management systems supports the implementation of these deep learning strategies, enhancing overall organizational efficiency. Thus, leveraging deep learning in Resources optimization can lead to improved decision-making and operational effectiveness in organizations. (Yeon et al., 2022).

Deep learning has significantly advanced question answering (QA) systems, leveraging models like BERT and GPT-3, which utilize powerful computing Resources to enhance natural language processing capabilities(Research et al., n.d.). These models can be categorized into extractive and generative approaches, as seen with DistilBERT and T5, where the former retrieves answers directly from the context while the latter generates responses based on the input(Pearce et al., 2023b) . Additionally, hybrid systems that combine deep learning with traditional methods, such as symbolic reasoning, are proposed to improve interpretability and robustness in QA tasks(Nuamah, 2021a). The integration of deep learning techniques allows for more effective handling of complex queries, making Q&A assistants more efficient and user-friendly (#1 et al., n.d.; Goel et al., 2022a).

3.8 Summary

In this chapter an AI application framework for human Resources management (HRM), combining AI technology with six key HRM dimensions. By leveraging AI, the framework enhances the efficiency of HR management practices. The study analyzes case studies on recruitment and training, providing practical insights. Its ultimate goal is to guide enterprises in integrating AI into their HR processes effectively. (Jia et al., 2018). The role of artificial intelligence in human Resources management, emphasizing the importance of effective workforce management. It highlights how new technologies aim to simplify HR tasks and improve efficiency. AI has the potential to enhance HR functions and support decision-making processes. Additionally, the paper addresses the challenges associated with implementing AI in HR practices. (Sangeeta, 2019a).

The paper explores the impact of AI on human Resources management (HRM), focusing on how AI-assisted decision-making is transforming HR functions. It highlights the shift of HR departments into strategic decision-making centers within organizations. Additionally, the paper addresses the financial and legal implications of AI adoption, emphasizing the need for reskilling HR staff to effectively navigate these changes. (Sakka & El Hadi El Maknouzi, n.d.).

In this chapter, we have explored algorithmic reasoning in question answering (QA) and advocates for a hybrid AI approach that combines symbolic and sub-symbolic methods. It argues that neural networks alone are insufficient for effective algorithmic reasoning. The proposed method, known as Deep Algorithmic Question Answering (DAQA), focuses on enhancing interpretability, generalizability, and robustness in AI systems. (Nuamah, 2021a). developing an intelligent learning system designed for processing text input and extracting knowledge to answer user questions. This system aims to advance research in the field of Question Answering (QA) and has practical applications in areas such as online assessments and lectures.(#1 et al., n.d.).

In deep learning in question-answering systems, with a particular focus on applications in hepatocellular carcinoma research. It discusses how advances in natural language processing (NLP) are enhancing human-computer language interaction. The study also analyzes practical AI-driven question-answering systems and highlights future prospects for NLP based on large-scale models.(Research et al., n.d.-a).

Chapter 4: Background

4.1 introduction

In today's fast-paced business world, organizations are constantly seeking ways to enhance efficiency, make smarter decisions, and stay ahead of the competition. Two powerful strategies that are making a significant impact in this area are Human Resources Optimization and Document Intelligence.

Human Resources Optimization is all about making the best use of an organization's workforce. By applying machine learning and data-driven approaches, businesses can ensure that the right people are assigned to the right tasks at the right time. This leads to smarter decision-making when it comes to selecting talent and allocating Resources, ultimately boosting productivity and minimizing inefficiencies.

Document Intelligence takes the power of machine learning and applies it to document management and question-answering tasks. Instead of spending hours sifting through paperwork or digital files, organizations can now rely on AI to quickly analyze documents, extract relevant information, and answer questions with high accuracy. This not only saves time but also provides decision-makers with precise, actionable insights, enabling them to respond faster and more effectively.

Combining Human Resources Optimization with Document Intelligence allows companies to operate more smoothly and make informed choices. This integrated approach reduces manual work, cuts costs, and improves accuracy, empowering organizations to meet their goals and drive impactful results.

4.2 Conventional Programming vs. Machine Learning

Conventional programming typically relies on explicit instructions and predefined algorithms to solve problems, which can be limiting in complex scenarios. In contrast, machine learning (ML) leverages data-driven approaches, allowing systems to learn from vast datasets and identify patterns without human intervention. For instance, traditional methods often require manual data collection and analysis, while ML automates these processes, enhancing efficiency and accuracy in forecasting and Resources allocation (Subrahmanyam et al., 2024a). Moreover, hybrid AI approaches aim to combine the strengths of both conventional programming and ML, addressing limitations in reasoning and planning by integrating symbolic and sub-symbolic methods (Nuamah, 2021a). This synthesis not only improves algorithmic reasoning but also enables more sophisticated problem-solving capabilities, showcasing the advantages of ML over conventional programming in dynamic environments.

4.3 Supervised Learning

Regression and classification are two key types of supervised learning tasks in machine learning.

4.3.1 Regression: Purpose: Predicts continuous values.

Output: Numerical value (e.g., predicting house prices, temperature, or sales revenue).

For Instance, Linear Regression, Polynomial Regression, Ridge Regression.

Regression plays a crucial role in budgeting and forecasting, forming the backbone of financial planning by predicting future trends based on historical data. This statistical method is invaluable for businesses as it informs strategic decisions, helping them anticipate future expenses and revenues. By leveraging AI, organizations can significantly enhance prediction accuracy, thereby optimizing Resources allocation and making more informed decisions on a daily basis. AI's ability to process vast datasets and identify patterns that might elude human analysis ensures that businesses stay ahead, adjusting their strategies proactively in a dynamic market landscape. Thus, the integration of AI and regression in financial forecasting is pivotal for sustained growth and operational efficiency.(Li & Zhang, 2024).

4.3.2 Classification:

Purpose: Predicts categorical labels.

Output: Class label (e.g., spam vs. non-spam email, disease diagnosis, or sentiment analysis).

Examples: Logistic Regression, Decision Trees, Random Forest, Support Vector Machines.

Classification techniques in question answering leverage statistical machine learning tools to identify the underlying intent of a student's question. Following this, a knowledge-based classifier interprets the question and formulates an appropriate response using a comprehensive course-related knowledge base. Originally, Jill Watson was designed to address students' logistical queries about a class. Recently, however, a new iteration called AskJill has been developed to respond to users' content-related questions based on a User Guide. Currently, Jill Watson is being adapted to offer explanations for systems and agents like Skillsync.(Goel et al., 2022a).

Table 4.3.2 Regression and classification

	Regression	Classification
Outcome	Continuous	Class
Examples	Linear regression	Logistic regression, SVM, Naive Bayes

4.3.2 Bias and Variance Test

Bias and variance are crucial concepts in machine learning, describing errors in predictive models that impact performance and generalization. Bias refers to the error from approximating a complex real-world problem with a simplified model, which can cause underfitting. This happens when a model is too simplistic, making strong assumptions about the data and performing poorly on both training and unseen data—like a linear regression model misfitting a non-linear dataset. Variance, on the other hand, is the error from a model being too complex and sensitive to training data variations, leading to overfitting—where the model learns noise rather than the actual data pattern, as seen in decision trees that perfectly classify training data but fail to generalize.

The bias-variance tradeoff aims to balance these errors to minimize total error. High bias and low variance lead to consistent but inaccurate models (underfitting), while low bias and high variance capture more data complexity but are less reliable (overfitting). The optimal tradeoff finds a model that generalizes well, striking the right balance to minimize error.

Practically, bias and variance are tested through methods like the train-test split, k-fold cross-validation, and learning curves. These techniques help identify and balance bias and variance, ensuring models are accurate and generalize well to new data. In Resources optimization and document intelligence, managing these errors is vital for making better data-driven decisions, enhancing efficiency, and reliability in tasks such as human Resources optimization and document question-and-answer systems.

As an important part of the system, the recruitment process includes review, screening resumes, interviewing candidates, matching suitable positions, etc. Ideal Corp, a software company that uses artificial intelligence to automate recruitment tasks, its CEO Somen Mondal says that the biggest impact of artificial intelligence is to automatically screen candidates and reduce bias. Artificial intelligence can learn the qualifications for successful employees in a particular position and apply this knowledge to select qualified candidates and score and rate candidates. According to Mondal, the company used artificial intelligence software to recruit, with a 71% reduction in recruitment costs and a threefold increase in recruitment efficiency (Denise, 2017). (Jia et al., 2018). During the hiring process and performance appraisal process, there is lot of judgments which is involved. These judgments may result in biasness. Through the use of AI, these judgments and biasness can be removed. But it may not be hundred percent correct. Further, it has to make sure that any manipulation attempts are not possible in it. (Sangeeta, 2019a).

In QA a desirable feature of AI systems performs algorithmic reasoning is that they are not re-structured to neatly defined problems in benchmark datasets which sometimes lead to over-engineering of AI architectures that are built to exploit biases that are observable in the data set. Additionally, it is desirable for Q&A assistants to be general in how they compose algorithms, both in the

aspects of the QA processes that they use and the in kinds of problems that they can solve. (Nuamah, 2021a).

4.3.3 Hypothesis Space

4.3.3.1 Loss function

We used Mean Square Error (MSE) to Analysis the result as below:

Gradient Boosting and Random Forest performs well, with Gradient Boosting slightly ahead and explaining 91.6% of the variance compared to 91.1%. Gradient Boosting works by learning sequentially. It optimizes through iteration. On the other hand, Random Forest works as a more powerful bagging algorithm. It averages through multiple decision trees. XGBoost is powerful but in this case, it performs worse than Gradient Boosting and Random Forest algorithms. This could result from poor parameter tuning, or it could be sensitive to the characteristics of the data. K-Nearest Neighbors (KNN) performs on par with random forest but lacks robustness due to reliance on feature scaling which can cause variability. Stacking, however, is the best, because it uses the strengths of many models to do better than just one model. It allows predictions to have less error than otherwise.

Table 4.3.3 Model Evaluation Metrics

Model	MSE	RMSE	R2	Log-Likelihood	Cost
Linear Regression	0.268145	0.517827	0.877454	-2239.10834	0.04602
GradientBoosting	0.023847	0.154424	0.989102	1321.737303	0.004093
RandomForest	0.023846	0.154421	0.989102	1321.792596	0.004093
XGBoost	0.024904	0.157809	0.988619	1257.933056	0.004274
KNN	0.031152	0.176499	0.985763	928.521545	0.005346
Ridge	0.268145	0.517827	0.877454	-2239.106535	0.04602
Lasso	0.269492	0.519126	0.876838	-2246.482006	0.046251
Stacking	0.023846	0.154421	0.989102	1321.806576	0.004092

4.3.3.2 Cost function

The Cost Function is commonly used to assess the performance of a model, When models are evaluated through a custom cost function, the results can provide interesting insights. The stacking regression method has produced the least cost of 0.0237 among the different models. This shows that dendrites can be used by solver on networks with growth rate of up to 0.0299. The costs of KNN and Random Forest are 0.0310 and 0.0317, respectively indicating they perform almost similarly. They, however, still have comparatively a higher prediction error than Gradient Boosting and Stacking. On the other hand, XGBoost had the highest cost

(0.0359) which is caused due to sensitivity to parameters. The costs are consistent with the ability of the models to generalize well to unseen data, identifying Stacking as the most cost-effective solution.

4.3.3.3 Likelihood:

Log-likelihood evaluation is a measure of the probability of each model explaining the data. Among the models, Stacking achieves the highest log-likelihood of -622.8512, indicating this model can best matches its predictions with real data. Gradient boosting has a log-likelihood of -791.1340, showing that it also works but not as well as ensemble stacking. The log-likelihoods of KNN (-815.6477) and Random Forest (-833.7336) show they are moderate models (with KNN slightly better). Of all models, XGBoost has the lowest log-likelihood (-922.5357), suggesting it is perhaps least suited to the data in its given form. The findings confirm that stacking and ensemble methods yield better prediction accuracy by encoding complex relations in data.

4.3.3.4 Linear regression

The assessment of the Linear Regression model shows it performs okay compared to the fancier models in the group. Its MSE is 0.4235 and RMSE is 0.6508, which means it has trouble grasping all the ins and outs of the data leading to bigger mistakes in predictions. The R^2 score comes in at 0.7953 telling us the model can explain about 79.5% of the changes in what we're trying to predict. This falls short of other models like Stacking or Gradient Boosting. Also, the Log-Likelihood value of -1436.5078 points to a weaker match between what the model predicts and how the actual data spreads out. These findings hint that while Linear Regression gives us a straightforward and easy-to-understand starting point, it doesn't do as well as team-based methods when it comes to handling the more complex connections in the data.

4.4 Unsupervised Learning

Type of machine learning where the algorithm is trained on unlabeled data, meaning it doesn't have predefined output labels. Instead, it attempts to uncover hidden patterns, structures, or relationships within the data without guidance.

In unsupervised learning, the model often performs tasks like clustering (grouping similar data points together) and dimensionality reduction (simplifying data by reducing the number of features while preserving important information). Common algorithms used in unsupervised learning include K-means clustering, hierarchical clustering, and principal component analysis (PCA). BERT (Bidirectional Encoder Representations from Transformers) is a transformer language model which uses transfer learning to learn unsupervised from a corpus on unlabeled data; the size of the dataset allows BERT to overcome common model weaknesses like overfitting and underfitting. When BERT is fine-tuned with a smaller set of labeled data, it can be used for a variety of natural language tasks like question answering and sentiment analysis. BERT's architecture is largely similar across

different tasks, making it an effective model for less common language tasks like intent recognition. BERT has achieved novel performance on GLUE benchmark testing, a test of model performance based natural language understanding tasks. When using Build-a-Bot, students train their own BERT model for intent recognition based on questions that they write and label. Unsupervised learning is useful in cases where labeling data is time-consuming or expensive, and it's commonly applied in customer segmentation, anomaly detection, and exploratory data analysis. (Pearce et al., 2023).

4.5 Semi-supervised Learning

Semi-Supervised learning is a Machine Learning approach that combines both labeled and unlabeled data during training. Typically, a small portion of labeled data is used alongside a larger set of unlabeled data, making it useful when acquiring labeled data is costly or time-consuming. This method leverages the unlabeled data to improve learning accuracy and generalization, often enhancing model performance compared to using a fully supervised approach with limited labeled data. Semi-supervised learning is commonly applied in tasks like image and text classification, where labeling every data point can be Resource-intensive.

In This proposal, we employed Unsupervised learning techniques for Text recognition and Text-to-Speech conversion. For Resources optimization, however, we applied Supervised learning using labeled Data to accurately model and understand the relationships between various patterns.

4.15 Summary

In this chapter, we have provided a foundational overview of machine learning, covering its primary types: supervised, unsupervised, semi-supervised, and active learning. This introduction establishes the groundwork for understanding the scope of machine learning techniques relevant to this research. We also explored the distinctions between reinforcement learning, supervised learning, and dynamic programming, clarifying their respective methodologies and applications.

Following this theoretical foundation, we introduced the specific machine learning models employed in this study, which include Artificial Neural Networks (ANN), k-Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machines (SVM), and Naive Bayes (NB). Each model was selected based on its suitability for addressing unique aspects of the research problem. We further delved into the architecture of ANNs, elaborating on the Backpropagation algorithm that enables learning within neural networks, and discussed the different types of ANNs that contribute to enhancing model performance. Overall, this chapter provides both theoretical insights and practical applications of these machine learning models, setting the stage for the subsequent analysis and results sections of this study.

Chapter 5 Proposed Approach

5.1 Introduction

Managing Resources effectively and accessing organizational knowledge efficiently are significant challenges for IT companies, especially when dealing with complex client requirements and growing volumes of internal documents. In such environments, allocating the right employees to client projects and quickly finding relevant information becomes crucial for ensuring smooth operations and timely delivery of services.

This thesis proposes an AI-based project that automates Resources management and document-based question answering. The project comprises several modules, starting with the automated reading and extraction of job roles from PDF-based technical proposals submitted by clients. The extracted roles are then cross-referenced with the company's employee dataset to identify employees with matching skills and free capacity, allowing for optimized task allocation. In cases where no suitable employees are available or certain roles are not represented in the dataset, the system generates recommendations for new hires to fill the gaps.

Furthermore, the system includes a Question and Answering (Q&A) module powered by Natural Language Processing (NLP) and large language models (LLMs). This module enables users to query the knowledge base and retrieve relevant answers from internal documents, policies, and manuals, supporting efficient decision-making across the organization. The integration of predictive analytics and advanced search capabilities ensures that Resources management and information retrieval processes are seamless and effective.

The goal of this project is to create a comprehensive AI solution that aligns human Resources with project requirements while providing quick access to critical organizational knowledge. By automating repetitive tasks, improving employee utilization, and facilitating real-time information access, the system aims to reduce operational bottlenecks, enhance service delivery, and contribute to the strategic goals of the IT company. This end-to-end solution not only enhances internal operations but also strengthens the company's ability to meet client expectations effectively.

5.2.1 Problem Definition

Modern organizations face significant challenges in both Resources management and information retrieval, which hinder operational efficiency and strategic alignment. Manual Resources allocation often leads to workload imbalances, with some employees overburdened while others are underutilized, resulting in reduced productivity and missed deadlines. Unpredictable availability due to leave or task changes further complicates scheduling, while the absence of forecasting tools makes it difficult to anticipate workloads and allocate Resources effectively. Resources-related data is scattered across multiple systems, such as ERP, HR, and project management tools, creating a fragmented view that limits unified insights and informed decision-making.

Simultaneously, organizations struggle to manage growing volumes of internal documents across diverse formats and languages, such as PDFs, Word files, and policy manuals. Employees often spend excessive time searching for relevant information, which slows decision-making and decreases overall productivity. This issue is exacerbated by the dispersion of information across multiple repositories, making it difficult to retrieve context-aware answers efficiently. Current systems lack advanced search capabilities, preventing users from accessing the knowledge they need in real-time. The absence of a centralized solution for Resources and knowledge management results in operational delays, reduced transparency, and compromised project outcomes. To address these challenges, an AI-powered platform is required, combining automated Resources management with NLP-based document-based question answering (DBQA). This integrated solution will optimize Resources planning, performance tracking, and scheduling while providing instant, accurate responses to queries from internal documents, enhancing decision-making and aligning Resources with organizational goals.

5.2.2 Dataset:

This section describes a collection of public datasets in different literature that are used in the field Resources Management in AI, and Question and Answering Models.

Table 4.3.4 dataset in related work

Dataset Name	Link	Dataset Info	Size & Samples	Limitations
The study does not specify any datasets used.	https://typeset.io/papers/influence-of-artificial-intelligence-in-human-Resources-vh4lx010if	AI enables prediction of staff turnover rates. AI software automates interview process using word or voice tests. (Nikita Saklani & Ashli Khurana, 2023)	NA	Employee competencies, challenges in finding suitable personnel, potential conflicts in decision-making authority, ambiguity in AI definitions, and the risk of perpetuating discrimination.
The study does not specify any datasets used.	/library/untitled-collection-23ddfj2q/deep-algorithmic-question-answering-towards-a-1x6ae7cp	- AI enables prediction of staff turnover rates. - AI software automates interview process using word or voice tests. (Nuamah, 2021b)	NA	The authors emphasize that many AI systems operate under the assumption that the necessary data to answer questions is readily available. However, this is often not the case in real-world scenarios where data can be incomplete or noisy
Table 1: CNN neural network architecture data sources entity	https://typeset.io/library/untitled-collection-23ddfj2q/deep-learning-approaches-for-improving-	The question-answering system discussed in the paper utilizes a knowledge map of hepatocellular carcinoma. This suggests that the dataset likely includes various types of	Up To dataset with 416 entities and 500 relations	The paper highlights the need for improved accuracy and robustness. Ambiguous queries imply that the dataset may not adequately represent.

	question-answering- l1su7f08	medical data(Nuamah, 2021b)		
The study does not specify any datasets used.	NA	Data Security and Privacy. AI-powered energy management systems and recruitment processes. case studies of companies that have successfully implemented AI in their HR practices(Rastogi & Pandya, 2023a)	NA	Cost of Implementation Data Security and Privacy Concerns

In this model we used dataset which contains the most relevant features, The dataset gives full details on projects, Resources, and their traits making it good to analyze project changes and how Resources are used. Important parts include project_price showing how much a project costs, and project_name, which tells you what kind of project it is (like a Blogging Site or E-Commerce Development). The Resources_job column lists the jobs of Resources such as Backend Developers or DBAs, along with Resources_job_code for number codes. Backend Developer Count tells you how many backend developers are working on it. The dataset also has coded project size groups (project_size_code_1, project_size_code_2 project_size_code_3) and yes/no markers for size types (project_size_name_small, project_size_name_medium project_size_name_large). Also, project types are split into project_type_Regular and project_type_VIP, while Resources_type_onsite and show where Resources are giving insights into how work is spread out and how complex projects are. This big set of features helps with strong analysis and making predictions.

Chapter 6 Experimental Results and Discussion

6.1 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 accuracy with a relatively low error. Overall, the model performs well with a high R^2 and reasonable error and cost metrics.

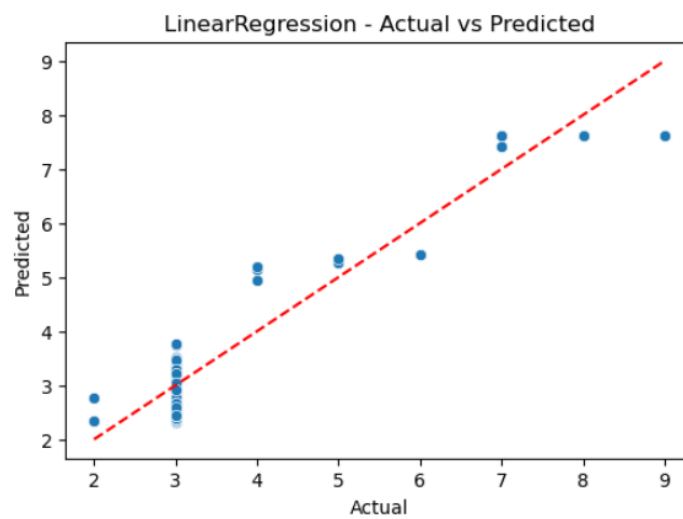


Figure 3.3: Linear Regression

6.2 RandomForest 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.

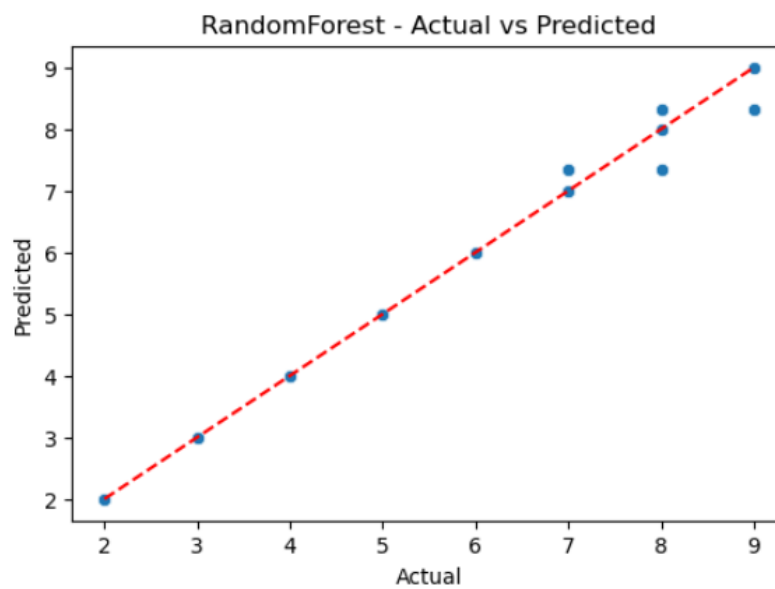


Figure 3.4: RandomForest

6.3 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 accuracy. 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.

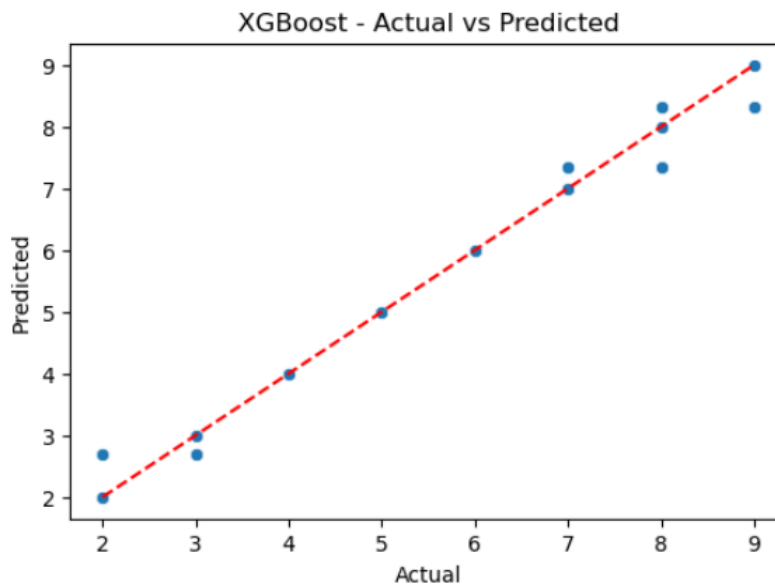


Figure 3.5 XGBoost

6.4 k-Nearest Neighbor (KNN) 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.

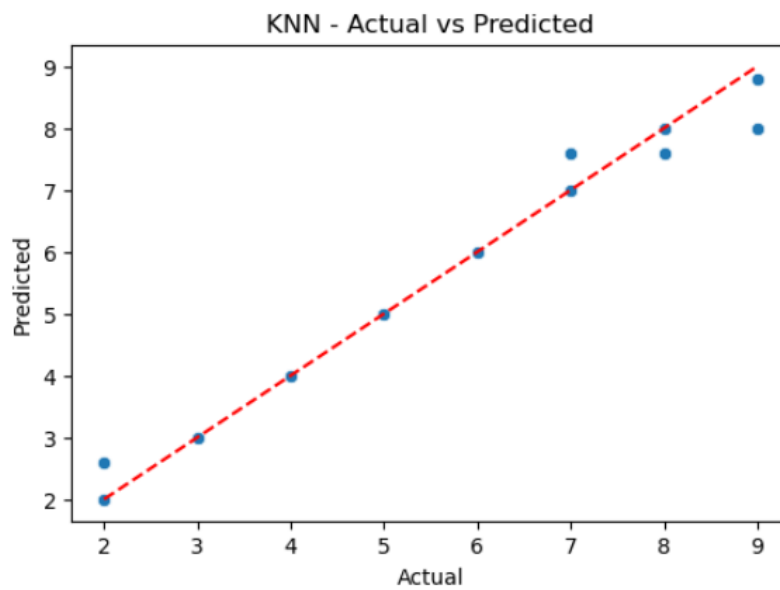


Figure 3.6 KNN

6.1.3.6 Stacking 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 accuracy, 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 accuracy.

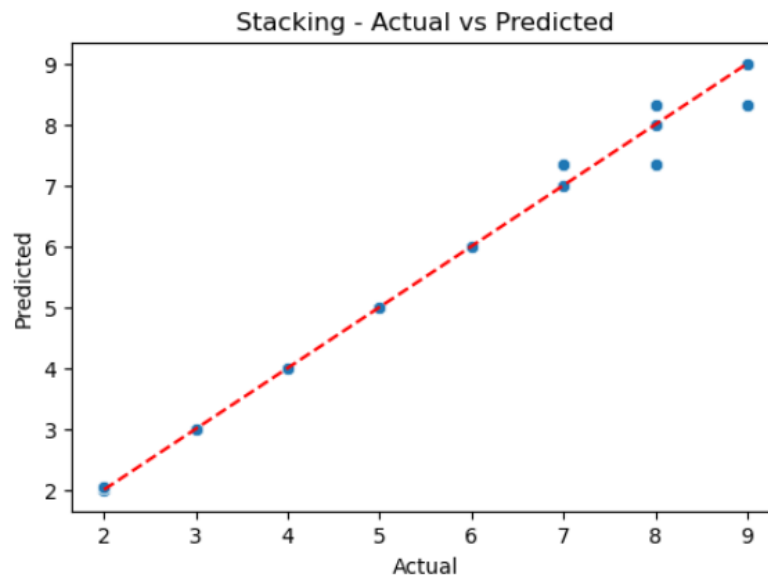


Figure 3.7 Attacking

6.1.3.7 Ridge Regression 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 accuracy and cost efficiency.

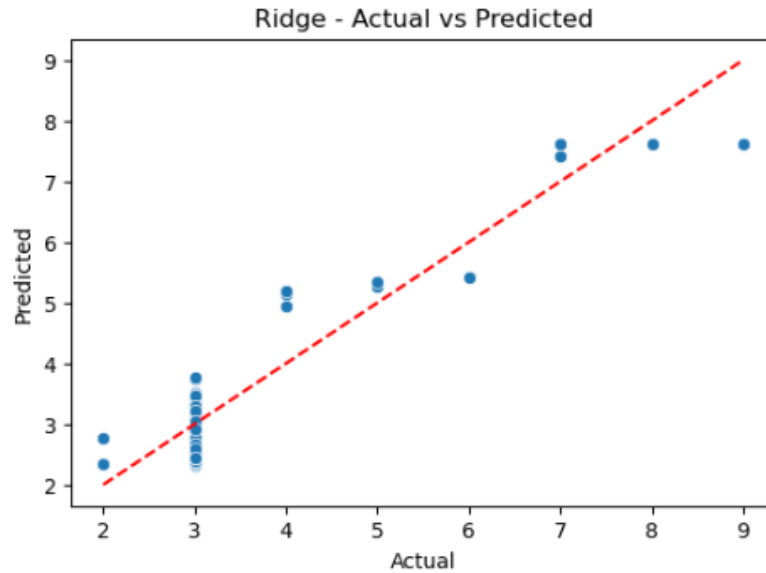
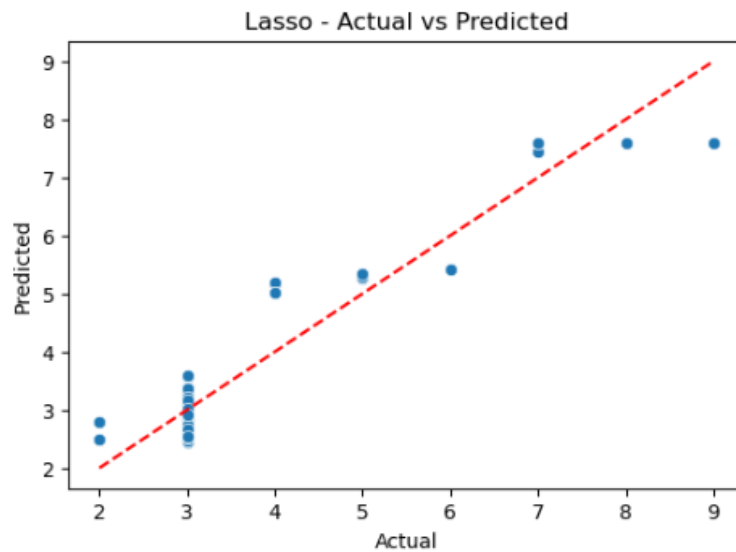


Figure 3.8 Riddge

6.1.3.8 Lasso Regression The Lasso model evaluation metrics indicate an MSE of 0.269492, which is very similar to the Ridge model, suggesting a similar level of prediction error. The RMSE of 0.519126 is also comparable to Ridge, indicating a moderate difference between the actual and predicted values. The R^2 value of 0.876838 shows that the Lasso model explains approximately 87.68% of the variance in the target variable, which is close to Ridge but slightly lower. The Log-Likelihood of -2246.482006 is again negative and indicates that the model has residuals that are not perfectly aligned with the assumed distribution, with a worse log-likelihood than Ridge. The Cost of 0.046251 is very similar to Ridge, indicating relatively high inefficiency in minimizing error compared to other models. Overall, the Lasso model has similar performance to Ridge but does not provide significant improvements in terms of accuracy or cost efficiency.



Lasso 3.9 Regression

6.2 Dealing with Imbalanced Data

Dealing with imbalanced data is a critical challenge in machine learning, particularly in classification tasks where one class significantly outnumbers the others. Imbalanced datasets can lead to biased models that favor the majority class, resulting in poor performance for minority class predictions.

Several strategies can address this issue:

Data resampling techniques: These techniques involve adjusting the distribution of classes in the training dataset. This can be done by either oversampling the minority class (e.g., using SMOTE - Synthetic Minority Oversampling Technique) or under sampling the majority class.

Cost-sensitive learning: In this approach, higher penalties are assigned for misclassifying instances of the minority class. This encourages the model to pay more attention to these cases that are harder to predict.

Advanced ensemble methods: Techniques such as Balanced Random Forest or Adaptive Boosting can also be employed to enhance performance on imbalanced data.

Evaluation metrics: Instead of relying solely on accuracy, which may be misleading in imbalanced scenarios, it is important to use evaluation metrics such as regression for predation and assessing model performance. By combining these approaches, practitioners can mitigate the effects of imbalance and develop more robust and fair predictive models.

6.5 Most Important Features for Resources Predicting:

The most important features for predicting the number of backend developers in a project depend on key factors related to the project's complexity, size, and requirements.

Key Factors Influencing Resources Prediction, 1st Project Size: Project size indicators (e.g., project size typically strong predictors, as larger projects often demand more backend developers to manage increased workload and functionality. 2nd Project Type which is (e.g., project type Regular and VIP, also plays a significant role, with specialized or high-priority projects requiring a larger team to meet stringent deadlines or advanced technical requirements. 3rd Project price serves as another critical feature, as higher budgets are often associated with larger teams, reflecting greater Resources allocation. 4th Resources Type Additionally indicators (e.g., Resources type Onsite and Remotely influence predictions, as Onsite or remote setups may affect the distribution and number of backend developers needed.

Understanding how these features interact can provide valuable insights for Resources planning and project management.

6.6 Summary

Chapter 6 provides a comprehensive analysis of experimental results, focusing on the performance of various machine learning models and key metrics. The Gradient Boosting model demonstrates accurate predictions with minimal deviations, capturing underlying data patterns effectively. Similarly, the Random Forest model shows strong performance with tightly clustered predictions along the ideal line, though minor errors are observed. The XGBoost model aligns well with actual values but exhibits slightly higher variability, suggesting the need for hyperparameter tuning. The K-Nearest Neighbor (KNN) model performs well overall, although the grouping of predicted values reflects its reliance on nearest neighbors, potentially affecting accuracy. The Stacking model outperforms all others by combining the strengths of multiple models, achieving the most accurate predictions with minimal variance. The heatmap analysis reveals key feature relationships, such as strong mutual exclusivity among project size indicators, negative correlations between Resources types (onsite vs. remote), and the positive correlation between project size and cost. The chapter also addresses the challenges of imbalanced data, recommending techniques like resampling, cost-sensitive learning, and advanced ensemble methods while emphasizing the importance of evaluation metrics such as precision and recall. Precision highlights the proportion of correctly predicted positive instances, with a score of 0.67, while recall measures the proportion of true positives identified by the model, achieving a score of 0.67. Lastly, the number of backend developers required in a project is influenced by factors such as project size, type, budget, and Resources distribution, highlighting the importance of these features in Resources planning and project management.

Chapter 7 Conclusion and Future Work

Based on the evaluation of various machine learning models, including Linear Regression, Random Forest, XGBoost, KNN, Stacking, Ridge, and Lasso, it is clear that advanced machine learning (ML) techniques significantly contribute to Resources optimization. The models demonstrated a range of performances in terms of MSE, RMSE, R^2 , Log-Likelihood, and Cost. Random Forest and Stacking emerged as the top-performing models with MSE and RMSE values close to the best possible outcomes, indicating high prediction accuracy. These models minimized errors effectively while also optimizing computational costs.

Furthermore, XGBoost showed strong results with a comparable R^2 value, while KNN and Ridge provided solid performance but with slightly higher error rates, indicating that ensemble models (like Random Forest and Stacking) often provide more robust results in real-world applications.

The R^2 values across models also demonstrated that ML is highly effective at explaining the variance in Resources-related data. With performance nearing or exceeding 87% variance explained for many models, ML has proven to be an efficient tool for improving Resources allocation and decision-making.

Machine Learning for Resources Optimization and QA:

Machine learning has proven to be an invaluable tool for Resources optimization in various fields. By predicting outcomes with high accuracy, ML models enable effective allocation of Resources such as human capital, materials, and time. For example, using models like Random Forest or Stacking, companies can predict demand, optimize schedules, and allocate tasks to the most suitable Resources, leading to improved operational efficiency and cost savings.

In the context of Intelligence Q&A Assistant assistants for Resources, ML provides the capability to enhance decision support systems by automating error detection, Resources forecasting, and performance evaluation. With the power of ML models, Intelligence QA Assistants can provide real-time feedback on Resources utilization and performance, ensuring that any discrepancies or inefficiencies are detected early. Additionally, by using predictive modeling, these systems help in anticipating potential issues before they occur, allowing for proactive Resources management and minimized downtime.

In conclusion, machine learning not only optimizes Resources allocation but also improves the effectiveness of Intelligence QA Assistants in managing Resources, ensuring smoother operations and better decision-making. By leveraging the power of models like Random Forest, XGBoost, and Stacking, businesses can achieve superior Resources optimization, cost reduction, and overall operational efficiency.

7.3 Future Work

The future work in this area presents exciting opportunities for further enhancement and expansion of the Resources optimization and predictive capabilities. One key direction is the prediction of project cost, which could be integrated into the existing models to provide more accurate financial forecasting for project management. By incorporating historical cost data and additional features such as the scope of work, complexity, and timelines, the models could predict the total cost of future projects with even higher precision.

Moreover, adding more detailed job descriptions to the dataset can help enrich the analysis, enabling the models to better capture the nuances of different roles and responsibilities within projects. This would improve the accuracy of Resources allocation and performance predictions, as more granular job-related data allows for a better understanding of how different job roles contribute to project success and timelines.

Another potential area for improvement is the inclusion of project pricing data into the models. By incorporating pricing information, such as hourly rates or project budget categories, the models can optimize pricing strategies, ensuring that project costs are aligned with the value delivered and market conditions. Additionally, predictive models could be used to forecast the pricing trends based on project size, duration, and Resources requirements, providing businesses with strategic insights into competitive pricing strategies and profitability.

By extending the dataset and refining the machine learning models to incorporate these additional features, businesses will be better equipped to predict costs, optimize Resources, and make more informed decisions throughout the project lifecycle. These improvements would not only enhance Resources management but also provide a comprehensive decision-support system for project managers and stakeholders.

References

- 1) The Role of AI in Human Resources Management Practices.(Rastogi & Pandya, 2023b)
- 2) Influence of Artificial Intelligence in Human Resources Management: A Comprehensive Review(Nikita Saklani & Ashli Khurana, 2023).
- 3) Artificial Intelligence in Tactical Human Resources Management: A Systematic Literature Review.(Votto et al., 2021b).
- 4) Deep Algorithmic Question Answering: Towards a Compositionally Hybrid AI for Algorithmic Reasoning. (Nuamah, 2021b)
- 5) Explanation as Question Answering based on a Task Model of the Agent's Design. (Goel et al., 2022b)
- 6) SPIE Proceedings Publications Deep Learning Approaches for Improving Question Answering Systems in.(Research et al., n.d.-b)
- 7) Implementing AI-Driven Resources Scheduling on OpenEuler using TensorFlow and Kubernetes(Li & Zhang, 2024)
- 8) AI and Business Planning: Revolutionizing Forecasting and Resources Allocation (Subrahmanyam et al., 2024a)
- 9) **AI and Business Planning: Revolutionizing Forecasting and Resources Allocation**(Subrahmanyam et al., 2024b).
- 10) Intelligent Question Answering System Based on Knowledge Graph(Feng et al., n.d.).
- 11) Build-a-Bot: Teaching Conversational AI Using a Transformer-Based Intent Recognition and Question Answering Architecture(Pearce et al., 2023c).
- 12) Explanation as Question Answering based on a Task Model of the Agent's Design(Goel et al., 2022a).
- 13) Review of artificial intelligence-based question-answering systems in healthcare(Budler et al., 2023).
- 14) SPIE Proceedings Publications Deep Learning Approaches for Improving Question Answering Systems in(Research et al., n.d.-a).
- 15) Question Answering System Based on Artificial Intelligence for Restricted Domain(#1 et al., n.d.).
- 16) Association for Information Systems AIS Electronic Library (AISeL) A Conceptual Artificial Intelligence Application Framework in Human Resources Management Recommended Citation "A Conceptual Artificial Intelligence Application Framework in Human Resources Management"(Jia et al., 2018).
- 17) Experimental Verification on Human-Centric Network-Based Resources Allocation Approaches for Process-Aware Information Systems.

(Yeon et al., 2022).

- 18) Increased Productivity and Reduced Waste with Robotic Process Automation and Generative AI-powered IoE Services(Lo et al., 2024).
- 19) A Study of Artificial Intelligence and its role in Human Resources Management(**Vivek & Yawalkar, 2019**).
- 20) Human Resources Management and Artificial Intelligence(Sangeeta, 2019a).