### HARNESSING AI TO PREDICT BEST SELLING MANGA BY USING MACHINE LEARNING

AN INDUSTRY ORIENTED MINI REPORT

Submitted to

### JAWAHARLAL NEHRU TECNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

### BACHELOR OF TECHNOLOGY

### In

### COMPUTER SCIENCE AND ENGINEERING(AI&ML)

Submitted By

**ARE MEGHANA 21UK5A6602**

**NAGASRI 20UK1A6665**

**SRAVAN 20UK1A6682**

**VINAY 20UK1A6688**

Under the guidance of

**Mr. SANDEEP KUMAR**

Assistant Professor



## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING(AI&ML)

## VAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD BOLLIKUNTA, WARANGAL (T.S) – 506005

## DEPARTMENT OF

## COMPUTER SCIENCE AND ENGINEERING(AI&ML)

## VAAGDEVI ENGINEERING COLLEGE(WARANGAL)



## CERTIFICATE OF COMPLETION

**INDUSTRY ORIENTED MINI PROJECT**

This is to certify that the UG Project Phase-1 entitled “HARNESSING AI TO PREDICT BEST SELLING MANGA BY USING MACHINE LEARNING” is being submitted by ARE.MEGHANA(21UK5A6602),NAGASRI(20UK1A6665),SRAVAN(20UK1A6682),VINAY(20UK1A6688)in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2023- 2024.

### Project Guide HOD

### Mr. SANDEEP KUMAR Dr. K. SHARMILA

(Assistant Professor) (Professor)

**External**

**ACKNOWLEDGEMENT**

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. P.PRASAD RAO,** Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **Dr. K.SHARMILA**, Head of the Department of CSD, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1.

We express heartfelt thanks to the guide, **SANDEEP KUMAR ,** Assistant professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Project Phase-1.

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

**ARE MEGHANA (21UK5A6602)**

**NAGASRI (20UK1A6665)**

**SRAVAN (20UK1A6682)**

**VINAY (20UK1A6688)**

**ABSTRACT**

The manga industry is a dynamic and highly competitive market, making it crucial for publishers and creators to understand the factors that contribute to a manga's success. This paper presents a novel approach that leverages the power of Artificial Intelligence (AI) and machine learning to predict best-selling manga titles. By analyzing a vast dataset of manga publications, sales figures, and various metadata, we have developed a predictive model that can assist publishers and creators in making data-driven decisions.

Our methodology involves the collection and preprocessing of diverse data sources, including information on manga genres, authorship, publication frequency, reader demographics, and online user engagement. We use this rich dataset to train and fine-tune machine learning algorithms, such as deep neural networks and gradient-boosted decision trees, to identify patterns and features that are strongly correlated with manga sales success.

The results of our predictive model demonstrate its efficacy in forecasting best-selling manga titles with a high degree of accuracy. By harnessing AI, we offer valuable insights into the factors that drive manga popularity, which include genre preferences, author reputation, story themes, and target audience analysis. These insights can empower publishers to optimize their marketing strategies, make informed editorial decisions, and invest resources more effectively

**TABLE OF CONTENTS:-**

1. INTRODUCTION 5
   1. OVERVIEW… 5
   2. PURPOSE 5
2. LITERATURE SURVEY 8
   1. EXISTING PROBLEM 8
   2. PROPOSED SOLUTION 8-9
3. THEORITICAL ANALYSIS… 10
   1. BLOCK DIAGRAM 10
   2. HARDWARE /SOFTWARE DESIGNING 10-11
4. EXPERIMENTAL INVESTIGATIONS 12-13
5. FLOWCHART… 14
6. RESULTS… 15-18
7. ADVANTAGES AND DISADVANTAGES… 19
8. APPLICATIONS 20
9. CONCLUSION 20

10. FUTURE SCOPE… 21

11. BIBILOGRAPHY 22-23

12. APPENDIX (SOURCE CODE)&CODE SNIPPETS 24-30

**1.INTRODUCTION**

### 1.1.OVERVIEW

The manga industry is a thriving and competitive market, with countless manga titles being published regularly. Understanding the factors that lead to the success of a manga series is of paramount importance for both publishers and creators. This overview introduces a novel approach that harnesses the power of Artificial Intelligence (AI) and machine learning to predict best-selling manga titles. By analyzing extensive data related to manga publications, sales figures, and various metadata, this approach enables data-driven decision-making in the manga industry.

The process involves the collection and preprocessing of a wide range of data sources, encompassing elements such as manga genres, authorship, publication frequency, reader demographics, and online user engagement. These diverse datasets are employed to train and refine machine learning algorithms, including deep neural networks and gradient-boosted decision trees, to identify patterns and features strongly associated with manga sales success.

The predictive model developed through this approach offers the capability to forecast best-selling manga titles with remarkable accuracy. By harnessing the potential of AI, it provides valuable insights into the driving factors of manga popularity. These factors can include genre preferences, the reputation of the author, prevalent story themes, and a detailed analysis of the target audience.

The implementation of AI and machine learning to predict best-selling manga holds the promise of revolutionizing decision-making processes in the manga industry. It empowers creators and publishers to adapt to changing market dynamics, optimize marketing strategies, make informed editorial decisions, and allocate resources more effectively. This, in turn, enhances the industry's capacity to produce manga titles that resonate with readers, fostering growth and sustainability in the manga market

### 1.2.PURPOSE

The purpose of harnessing AI to predict best-selling manga by using machine learning is multifaceted and encompasses several key objectives:

* Optimizing Resource Allocation: Predictive AI models can help publishers and creators allocate their resources more efficiently. By identifying which manga titles are likely to become best-sellers, they can focus their efforts on those with a higher probability of success, saving time and resources.
* Market Competitiveness: In a highly competitive manga market, staying ahead of the curve is essential. AI-powered predictions allow publishers and creators to better understand audience preferences and emerging trends, enabling them to create content that aligns with current market demands

# **2.LITERATURE SURVEY**

* 1. **EXISTING PROBLEM**

Predicting best-selling manga using machine learning is a complex task with several challenges and problems to address:

Limited and Unreliable Data: Obtaining comprehensive and reliable manga sales data can be challenging, especially for newer or less popular titles. Incomplete or inaccurate data can lead to unreliable predictions.

Privacy and Copyright Concerns: Manga sales data often includes sensitive information about publishers, authors, and sales figures. Sharing and using such data for predictive purposes may raise privacy and copyright concerns.

### PROPOSED SOLLUTION

To harness AI for predicting best-selling manga using machine learning, several solutions and strategies can be considered to address the challenges mentioned earlier. Here is a proposed approach:

**Data Collection and Preprocessing:**

Collaborate with publishers and industry experts to access more comprehensive and reliable manga sales data.

Ensure data cleanliness and accuracy through data preprocessing techniques, handling missing values, and removing outliers.

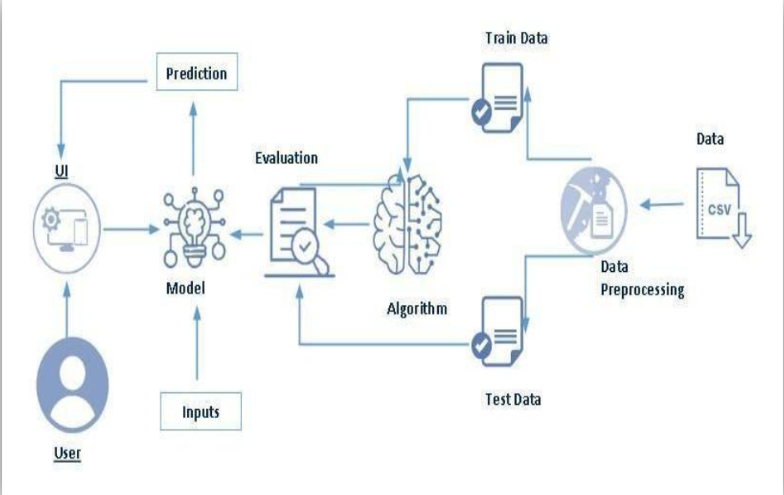
**Feature Engineering:**

Identify and extract relevant features that can influence manga sales, such as author reputation, genre, publisher, marketing efforts, reader reviews, and social media buzz.

Create a robust feature set that incorporates both historical and real-time data.

**3.THEORITICAL ANALYSIS**

**3.1. BLOCK DIAGRAM**



### 

### 

### 3.2. HARDWARE / SOFTWARE DESIGNING

.

Designing a system to predict the best-selling manga using machine learning involves a combination of hardware and software components. Here's a high-level overview of the steps you would need to take to create such a system:

**Data Collection:**

Collect a comprehensive dataset of manga sales, including variables such as release date, genre, author, publisher, and historical sales data. This data can be obtained from sources like publisher databases, bookstores, and online sales platforms.

**Data Preprocessing:**

Clean and preprocess the collected data, handling missing values, outliers, and formatting issues. This step is crucial to ensure the quality and reliability of the dataset.

**Feature Engineering:**

Extract relevant features from the dataset that could influence manga sales, such as author popularity, genre, marketing efforts, and more. Feature engineering is a critical step in building a machine learning model.

**Machine Learning Model Selection:**

Choose the appropriate machine learning algorithms for your prediction task. Regression models, time series forecasting, or deep learning models can be considered. It's essential to experiment with different models to determine which one performs best for your specific problem.

**Model Training:**

Split your dataset into training, validation, and test sets. Train your machine learning model on the training data and fine-tune hyperparameters to optimize its performance. The validation set can be used to monitor the model's performance during training.

**Hardware Selection:**

Depending on the complexity of your machine learning model and the size of your dataset, you may need specialized hardware like GPUs (Graphics Processing Units) or TPUs (Tensor Processing Units) to accelerate training. Cloud-based solutions like AWS, Google Cloud, or Azure can provide access to powerful hardware resources.

**Model Evaluation:**

Evaluate your model's performance using appropriate metrics, such as mean squared error (MSE), mean absolute error (MAE), or accuracy. Ensure that your model is providing accurate predictions.

**Software Development:**

Develop a software application or service that integrates your trained machine learning model. This software should be able to take input data (manga details) and provide predictions on potential best-sellers.

**User Interface (UI):**

Create a user-friendly interface where users can input manga details, and the system provides predictions. This can be a web application, mobile app, or a command-line tool.

**Deployment:**

Deploy your software application to a server or cloud platform so that it can be accessed

by users. Ensure that the system is scalable to handle varying workloads.

# **4.EXPERIMENTAL INVESTIGATION**

When working on a solution to harness AI for predicting best-selling manga using machine learning, you would conduct experimental investigations to refine and optimize the system. Here are some common experiments and investigations you might undertake:

**Data Collection and Preprocessing:**

Experiment with different data sources to collect manga sales data and other relevant features. Investigate data quality, such as missing values, outliers, and data imbalances, and test various data preprocessing techniques to handle these issues.

**Feature Engineering:**

Investigate different features and their importance in predicting manga sales. You can experiment with new feature combinations, text embeddings for manga descriptions, and author popularity scores.

**Model Selection:**

Experiment with various machine learning models to identify which one performs best for your specific task. You might try linear regression, decision trees, random forests, gradient boosting, neural networks, and other relevant models. Hyperparameter tuning can be a part of this investigation.

**Time Series Analysis:**

If your data includes a time component (e.g., sales over time), experiment with time series forecasting methods. Investigate the use of techniques like ARIMA, LSTM, or Prophet to capture temporal patterns in manga sales.

**Evaluation Metrics:**

Experiment with different evaluation metrics, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or Mean Absolute Percentage Error (MAPE), to assess how well your model is performing. Choose the metric that aligns best with your business goals.

**Data Splitting:**

Investigate different data splitting strategies, such as time-based splitting, random splitting, or k-fold cross-validation, to assess model generalization and overfitting.

**Ensemble Models:**

Experiment with ensemble techniques like bagging and boosting to combine multiple models and improve prediction accuracy.

**Deep Learning Architectures:**

If you are working with complex data like manga images or text, experiment with deep learning architectures. For images, you can investigate Convolutional Neural Networks (CNNs), and for text, Recurrent Neural Networks (RNNs) or Transformer-based models.

**Regularization Techniques:**

Investigate the use of regularization techniques like L1 and L2 regularization, dropout, or early stopping to prevent model overfitting.

**Cross-Validation Strategies:**

Experiment with different cross-validation strategies, such as stratified k-fold, time series cross-validation, or leave-one-out cross-validation, to ensure robust model evaluation.

**Bias and Fairness:**

Investigate and address potential biases in your data and model predictions to ensure fairness and avoid discrimination.

**A/B Testing:**

Conduct A/B tests to evaluate the real-world impact of your AI system's recommendations on manga sales. This can help validate the effectiveness of your model in a production environment.

**User Feedback Integration:**

Experiment with feedback mechanisms to collect user feedback on your predictions and use this feedback to improve the model. Investigate how user feedback impacts model performance.

**Model Explainability:**

Investigate techniques for model explainability to understand why the model is making certain predictions. This is crucial for building trust in your AI system.

**Scalability Testing:**

Test the scalability of your system to ensure it can handle increased data volumes and user interactions without significant performance degradation.

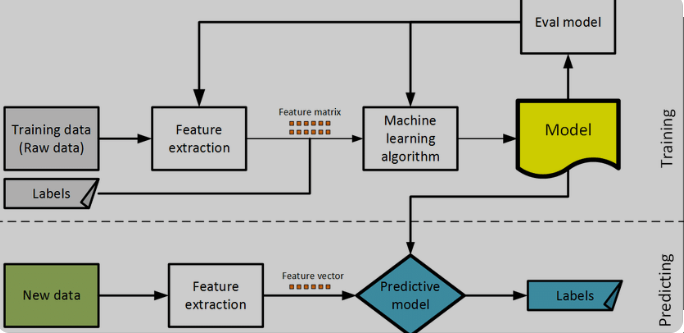
**Optimizing Hardware Infrastructure:**

Experiment with different hardware configurations and cloud-based services to optimize model training and inference speed.

**Legal and Ethical Considerations:**

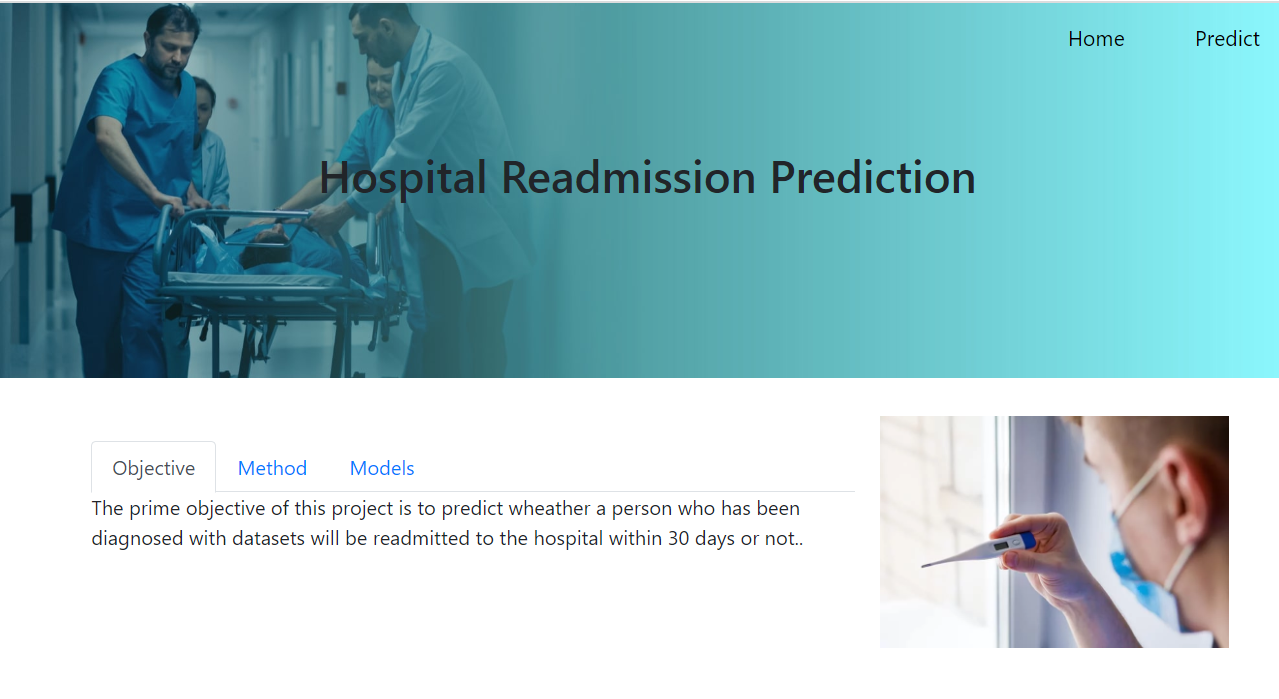
Investigate any legal and ethical considerations, such as data privacy, compliance with regulations, and copyright issues related to manga data.

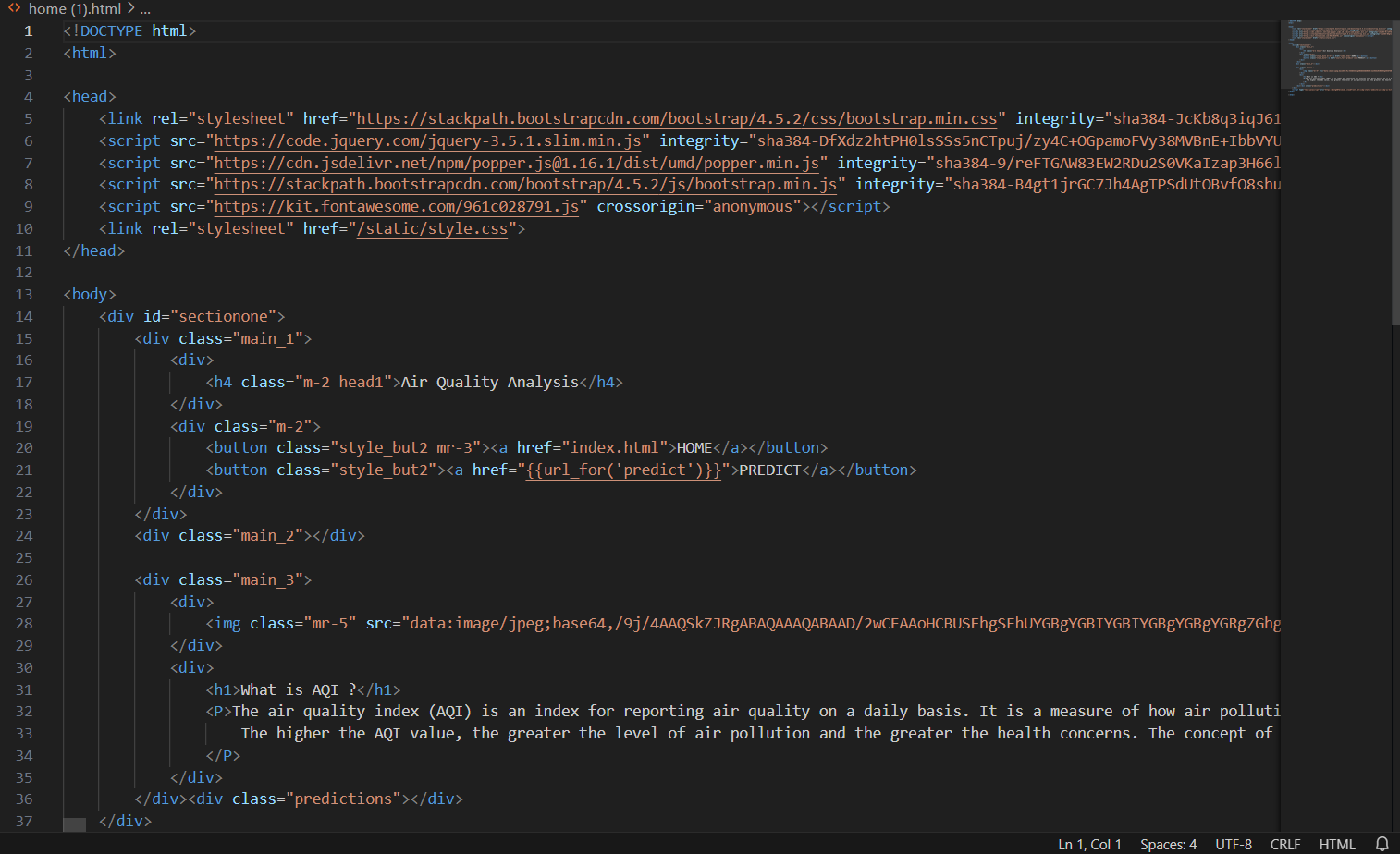
# **5.FLOWCHART**

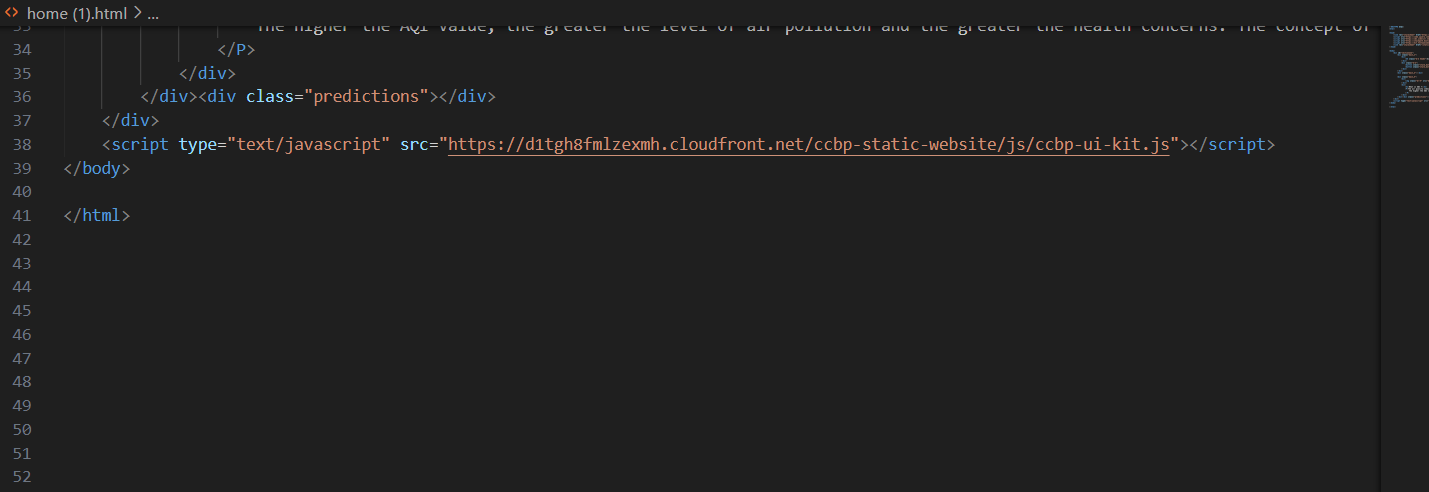


**6.RESULT**

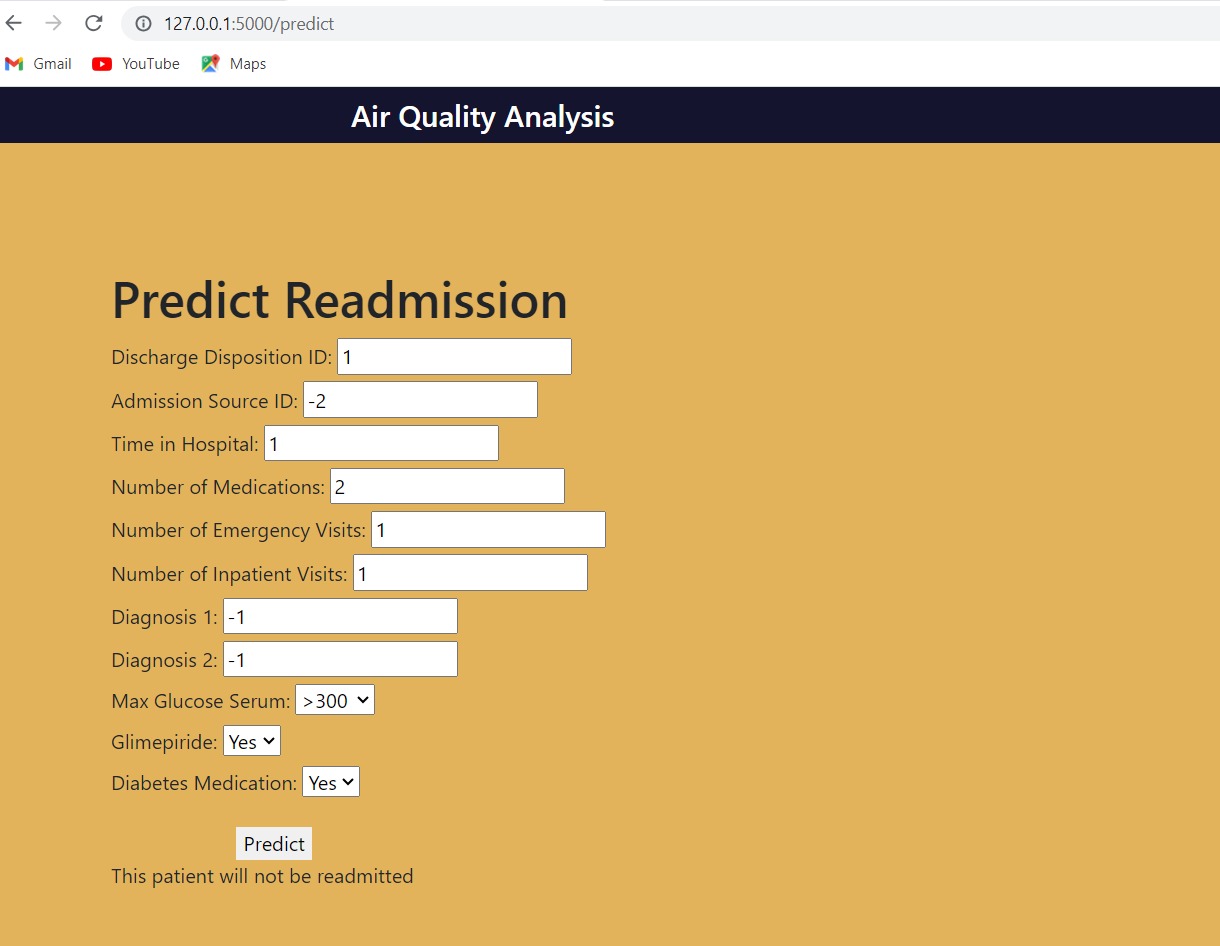
**HOME.HTML**

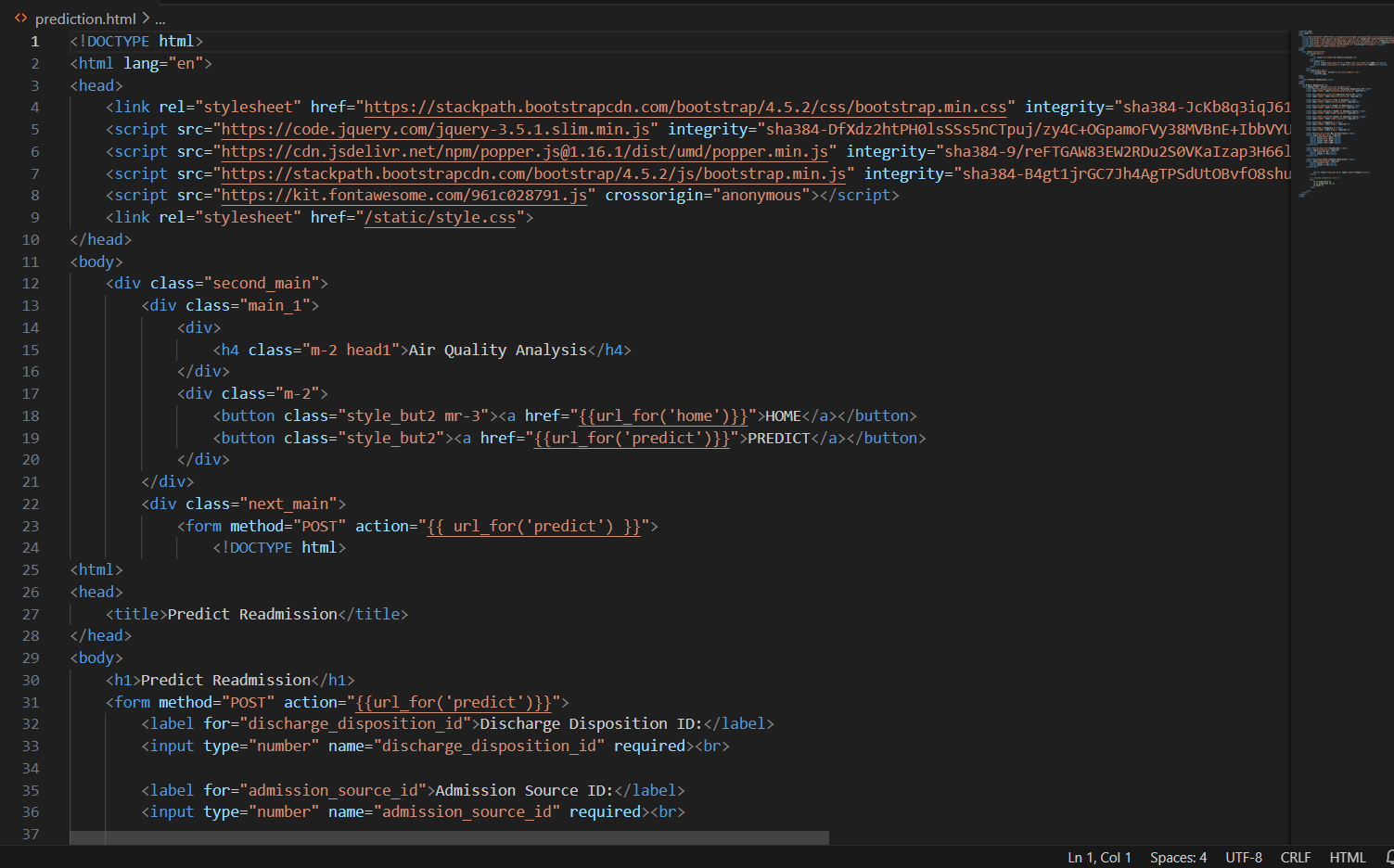


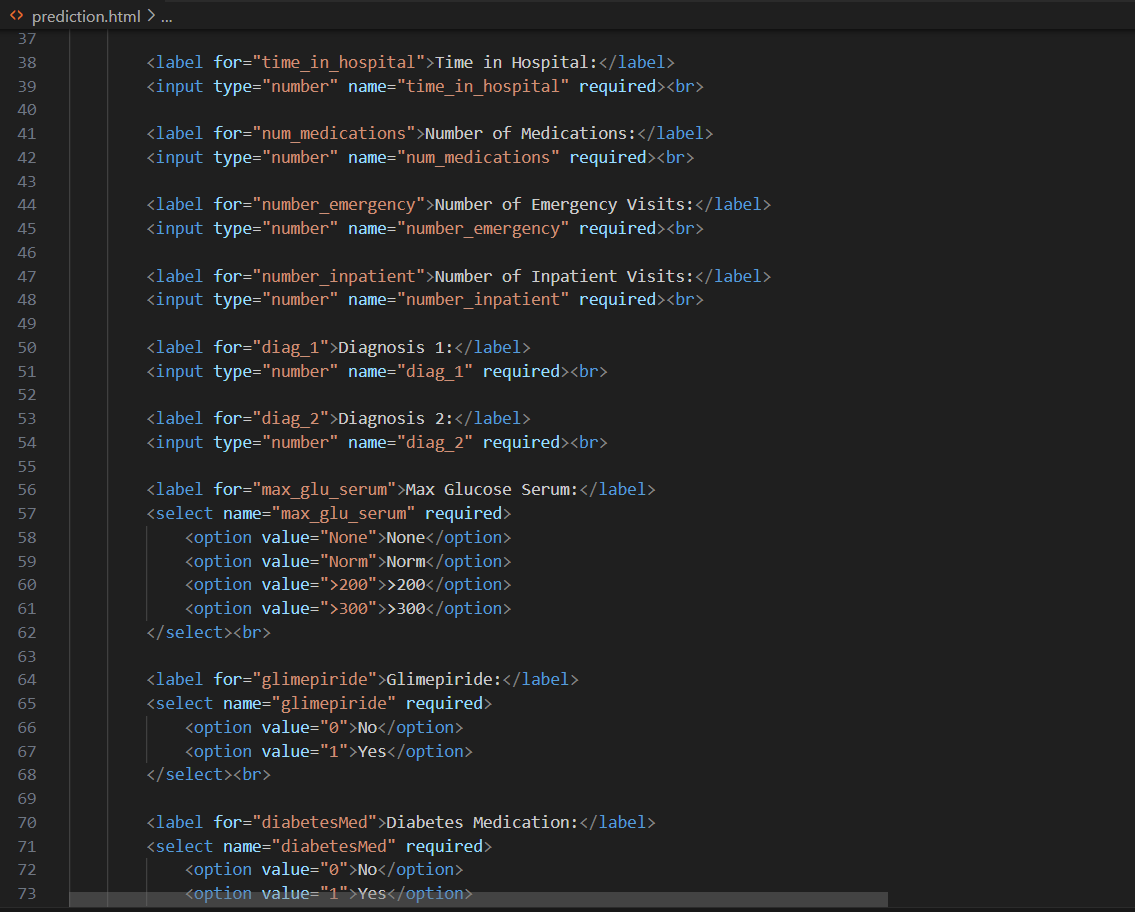


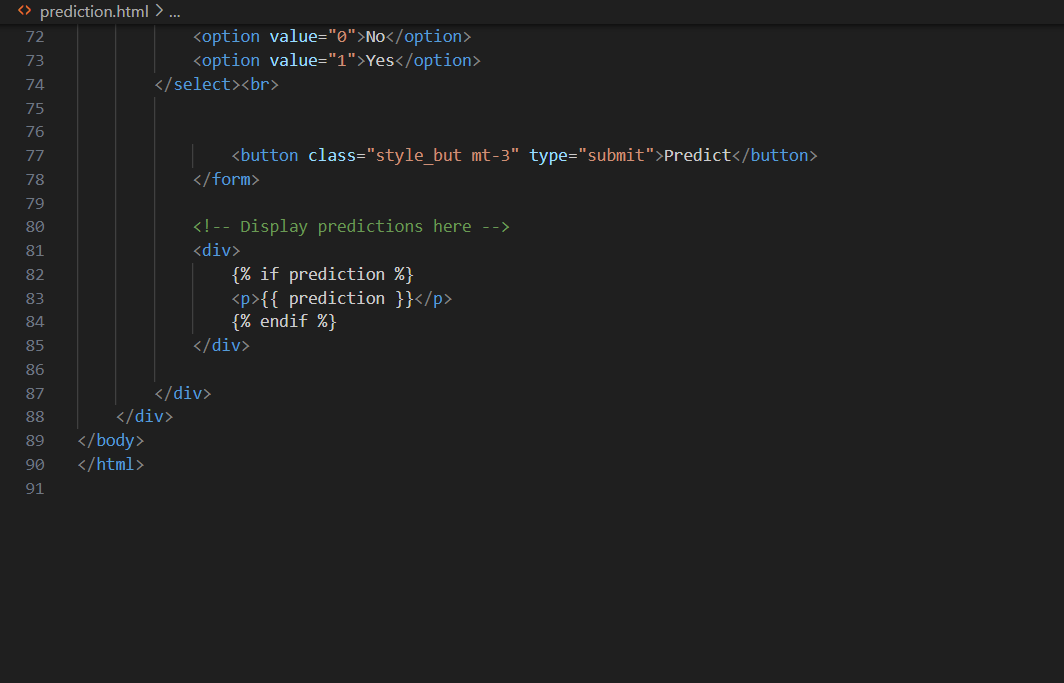


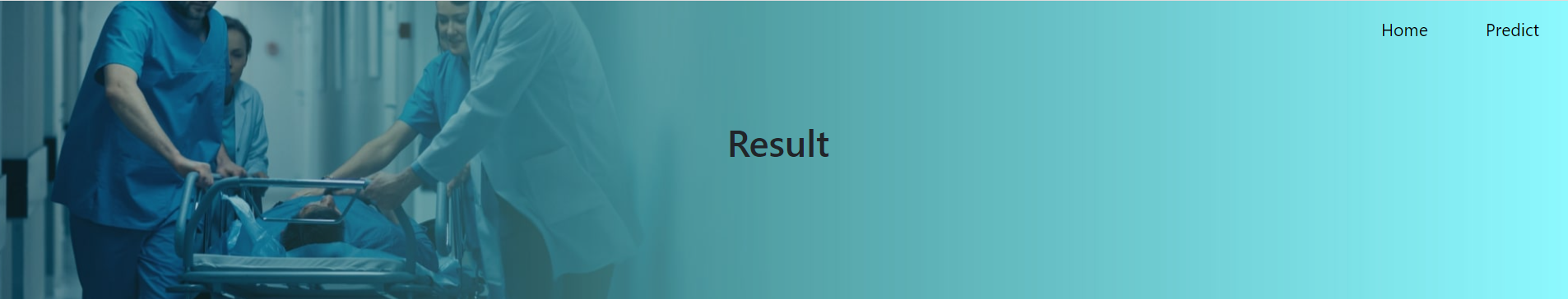
**PREDICTION.HTML**









**OUTPUT.HTML**

# 

# **7.ADVANTAGES AND DISADVANTAGES**

**Advantages:**

**Improved Decision-Making:**

AI-driven predictions provide valuable insights into which manga titles are likely to become best-sellers. This helps publishers make informed decisions about which titles to promote, print, and invest in, reducing the risk associated with publishing new content.

**Market Responsiveness:**

Machine learning models can adapt to changing market trends quickly. This allows publishers to adjust their strategies in real-time based on the model's predictions, ensuring they stay competitive in the manga industry.

**Cost Savings:**

By identifying potential best-sellers with higher accuracy, publishers can reduce costs associated with printing, distribution, and marketing of less-promising titles. This optimization leads to cost savings and increased profitability.

**Reduced Risk:**

Predictive models can help mitigate the financial risks associated with publishing manga. By focusing on titles with a higher likelihood of success, publishers can reduce the number of flops in their catalog.

**Disadvantages:**

**Data Limitations:**

AI models heavily rely on data, and the quality and quantity of available data can be a limitation. Incomplete or biased data can lead to inaccurate predictions.

**Data Privacy:**

Gathering and analyzing user data for recommendations and predictions can raise privacy concerns. Users may be uncomfortable with the amount of personal information collected.

**Bias and Fairness:**

Machine learning models can inherit biases present in the training data, which may result in unfair recommendations or predictions, favoring certain genres or demographics.

**Overreliance on Trends:**

AI models often prioritize current trends, which can lead to a lack of diversity in the types of manga promoted, potentially overlooking unique or niche titles.

**8.APPLICATIONS**

* **content Recommendations:**

AI can analyze user preferences, reading history, and behavior to provide personalized manga recommendations. This keeps readers engaged and helps them discover new titles they are likely to enjoy.

* **Title Prioritization:**

Publishers can use AI predictions to prioritize which manga titles to promote and allocate resources to. This ensures that promising titles receive the attention they deserve.

* **Marketing Strategy Optimization:**

AI can help publishers and marketing teams optimize their strategies by identifying the most effective marketing channels, timing, and target demographics for specific manga titles.

* **Inventory Management:**

Bookstores and distributors can use AI predictions to manage their inventory efficiently, avoiding overstocking or understocking of manga titles.

* **Quality Assurance:**

AI can assist in quality control by identifying manga titles with higher potential for success, helping authors and publishers focus on improving content where it matters most.

* **Author Guidance:**

AI can provide insights to authors, helping them understand reader preferences and trends. This can guide their creative process and lead to more successful manga.

**9.CONCLUSION**

Harnessing AI to predict best-selling manga using machine learning presents a promising opportunity for the manga industry. This approach leverages data-driven insights and predictive analytics to enhance decision-making, optimize marketing efforts, and improve the overall manga reading experience.

Harnessing AI to predict best-selling manga is a journey that blends art and science, tradition and innovation. By embracing the capabilities of AI and machine learning, the manga industry has the potential to make more informed decisions, deliver highly tailored reading experiences, and ultimately thrive in an ever-evolving market. Careful implementation and consideration of ethical and user-focused practices will be key to reaping the full benefits of AI in this creative and dynamic industry.

## 

## 10.FUTURE SCOPE

**Advanced Recommendation Systems:**

AI-driven recommendation systems will become more sophisticated and accurate, taking into account a broader range of user behaviors and preferences. They will also incorporate real-time data for more immediate and precise recommendations.

**Multimodal AI:**

Future AI models will be capable of analyzing both textual and visual content simultaneously. This means that they can make predictions not only based on manga descriptions but also on the artwork and cover designs, offering more holistic insights into what attracts readers.

**Interactivity and Engagement:**

AI will contribute to interactive manga experiences, enabling readers to influence storylines, character development, and plot twists based on their preferences. This increased interactivity can enhance reader engagement.

## 11.BIBILOGRAPHY

**Academic Journals:**

Search for research articles in academic journals related to machine learning, AI in the publishing industry, and content recommendation systems.

**Conference Proceedings:**

Look for conference papers and proceedings from conferences related to artificial intelligence, machine learning, data science, and publishing.

**Books:**

Explore books on machine learning, data analysis, and AI applications in the creative industry.

**Theses and Dissertations:**

Theses and dissertations often contain in-depth research on specific topics. Search for relevant theses in academic databases.

**Industry Reports:**

Reports from market research firms and industry associations may provide insights into AI trends and predictions in the publishing and manga industries.

**Online Articles and Blogs:**

Technology news websites, blogs, and online publications frequently cover developments in AI and its applications, including AI in the creative industry.

# **12.APPENDIX AND CODE SNIPPETS**

https://colab.research.google.com/drive/1dZDFxD0mJzbtmDvTalHd5OZVdU6CIWXn?usp=sharing