IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105174|

Wine Quality Prediction using Elastic Net

Sangam Adhikari, Asmita Gurung, Arpan Tuladhar, Emranul Haque Raque, Sumit Ghosh, Md Ahsanul Karim Likhon, Nishant Banjade, Vijay Kumar S

UG Student, Dept. of CSE., SET, Jain University, Bangalore, India

UG Student, Dept. of CSE., SET, Jain University, Bangalore, India

UG Student, Dept. of CSE., SET, Jain University, Bangalore, India

UG Student, Dept. of CSE., SET, Jain University, Bangalore, India

UG Student, Dept. of CSE., SET, Jain University, Bangalore, India

UG Student, Dept. of CSE., SET, Jain University, Bangalore, India

UG Student, Dept. of CSE., SET, Jain University, Bangalore, India

Assistant Professor, Dept. of CSE., SET, Jain University, Bangalore, India

ABSTRACT: Predicting wine quality is crucial for the wine industry and consumers. Wine quality determines safety for consumption and meeting connoisseurs' expectations. This study aims to predict wine quality by analyzing chemical and physical properties using statistics and machine learning. The Elasticnet algorithm addresses overfitting and underfitting. Models were trained and tested on a wine quality dataset with properties like pH, alcohol content, and density.

The models were integrated into a Flask web application using DevOps tools (pipelines, YAML, JSON) for continuous integration and deployment. The frontend (HTML, CSS, JavaScript) takes user input, applies it to the trained model's predict function, and displays the predicted wine quality value.

The web application provides insights for manufacturers and consumers by predicting values between 6 and 7. The research shows machine learning's potential in the wine industry and a foundation for further work. It highlights the importance of MLOps for automating and deploying models. Elasticnet effectively predicts wine quality, and the web application provides valuable insights.

KEYWORDS: Elastic Net, HTML, CSS, YAML, JSON

I. INTRODUCTION

Wine quality prediction is a complex task that involves the analysis of various factors such as grape variety, climate, soil, and production techniques. One of the popular approaches forpredicting wine quality is using machine learning algorithms. Among these algorithms, Elastic Net is a powerful and efficient method that combines both L1 and L2 i.e., Ridge and Lasso regularization to produce a model that is robust to outliers and can handle high-dimensional data.

Predicting the wine quality helps in objectively evaluating the quality of different wines. Wine producers can benefit from quality prediction models by optimizing their production processes. Winemakers can choose the production variables in a way that is guided by their analysis of the elements that go into producing high-quality wines. This may result in more control over uniformity and quality in the manufacturing of wine.

Elastic Net regression may handle multicollinearity (correlations between predictor variables) better than using either L1 or L2 alone by integrating the two regularizations. It maintains the stability and regularization effects of Ridge regression as well as the feature selection capabilities of Lasso regression. Since there are numerous correlated features or when working with high-dimensional datasets, Elastic Net regression is especially helpful in these situations.

DevOps plays a crucial role in ensuring the seamless development, deployment, and management of machine learning models. DevOps methodologies can help in automating the entire machine learning pipeline, including data pre-processing, model training, hyperparameter tuning, and deployment. This can significantly reduce the time and effort required to develop and deploy machine learning models.

Therefore, in this project, we aim to predict wine quality using Elastic Net algorithm and leverage DevOps methodologies to build a scalable and efficient machine learning pipeline. We will explore various data pre-processing

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105174|

techniques, hyperparameter tuning methods, and model evaluation strategies to build a robust and accurate wine quality prediction model.

II. RELATED WORK

[1]**A machine learning application in wine quality prediction**
Authors: Piyush Bhardwaj, Parul Tiwari, Kenneth Olejar Jr, Wendy Parr, and Don Kulasiri
Year Of Publication: 2022
Summary: This paper is our base paper. Datasets are collected from New Zealand from different regions of wine. This paper used different machine learning such as XGB, RF, GNB, AdaBoost, SGD, SVM, DTC and KNN and among them, AdaBoost model had high accuracy than other.

[2]**Wine Quality Prediction using Machine Learning Algorithms**
Authors: Devika Pawar, Aakanksha Mahajan and Sachin Bhoithe
Year Of Publication: 2019
Summary: Datasets are from Kaggle website of red wine. The author used classification algorithms such as Random Forest, Stochastic Gradient Descent, Support Vector Machine and Logistic Regression. First of all, data is imported and EDA and also data pre-processing is done. Then, it is splitted into training and testing and model are trained. The random forest has higher accuracy than compare to another model.

[3] **Prediction of Wine Quality Using Machine Learning Algorithms**
Authors: K. R. Dahal, J. N. Dahal, H. Banjade, and S. Gaire
Year Of Publication: 2021
Summary: Datasets are from Kaggle, UCL and machine learning repository of red and white wine. This paper compared the performance of Ridge Regression (RR), Support Vector Machine (SVM), Gradient Boosting Regressor (GBR), and multi-layer Artificial Neural Network (ANN) in which gradient boosting performed best to predict the wine quality.

[4] **Prediction of Quality for Different Type of Wine based on Different Feature Sets Using Supervised Machine Learning Techniques**
Authors: SatyabrataAich, Ahmed Abdulhakim Al-Absi, Kueh Lee Hui, and Mangal Sain
Year Of Publication: 2018
Summary: Datasets are collected from UCI database which have two variants of wine i.e., red and white wine. This paper mentioned about potential of genetic algorithm as well as simulated annealing algorithm for feature selection as well as the potentials of the classifiers to predict accurately based on the new feature sets.

[5] **Red Wine Quality Prediction Using Machine Learning Techniques**
Authors: Sunny Kumar, Kanika Agrawal, and Nelshan Mandan
Year Of Publication: 2021
Summary: Datasets are collected from UCI machine learning repository of red wine. This paper used the different techniques such as Random Forest, Support Vector Machine and Naïve Bayes on RStudio software and also calculated confusion matrix for each technique. Among them, SVM have highest accuracy than other techniques.

[6] **Prediction of Wine Quality Using Machine Learning**
Authors: T M Geethanjali, Sowjanya, M Y Rohith S N, Shubhashree B E and ShourishCharan A
Year Of Publication: 2021
Summary: The two datasets are related to red and white variants which are obtained by Kaggle. This paper used the different techniques such as Logistic Regression, Decision Tree Classifier, Random Forest Classifier, and Extra Tree Classifier for building the model and also calculated confusion matrix for each technique. Among them, Extra Tree Classifier have highest accuracy than other techniques.

[7] **Selection of important features and predicting wine quality using machine learning techniques**
Authors: Yogesh Gupta
Year Of Publication: 2018
Summary: The two datasets are related to red and white variants which are obtained by Kaggle. This paper used the different techniques such as Linear regression, neural network and support vector machine for building the model and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105174|

also calculated R square and Adjusted R square for each technique. Among them, SVM have highest accuracy than other techniques.

[8] **Red wine quality prediction through active learning**
Authors: Zhou Tingwei
Year Of Publication: 2021
Summary: The dataset used in this experiment contains several important indicators of 1600 bottles of red wine, as well as the quality assessed by famous wine tasters (from 0 to 10 stars). First of all, author did PCA then dataset is splitted into training and unlabeled pool and use KNN in unlabeled pool. Then, he used ranked batch-mode sampling to update the model and accuracy is calculated from every batch and this way accuracy is improved.

III. PROPOSED ALGORITHM

A. *Design Considerations:*
Our model consists of a user application based model,where a user interface connects a deep rooted backend consisting of NLP techiniques.
This system performs the basic exploratory data analysis and text preprocessing required for NLP. It has the ability to create models through a web-based graphical user interface. It just requires a dataset as input, and our web GUI outputs a dataset based on the user's option of word or phrase analysis. The user does not need any prior coding knowledge.

B. *Sequence Diagram*

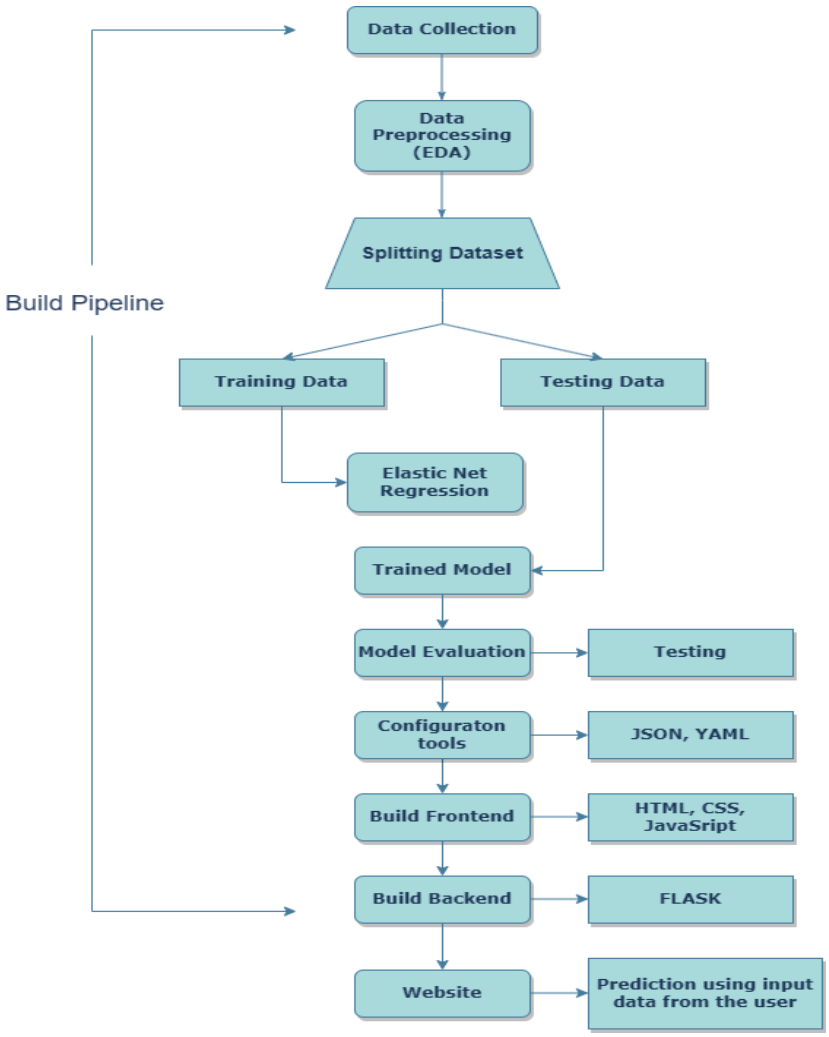

Fig: Flowchart of our proposed model.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105174|

C. *Description of the Proposed Algorithm:*

Data:
Firstly, all the necessary libraries that are required were imported and a dataset of wine data has been taken from Kaggle, where acidity, residual sugar, chloride, density, pH value, sulphates, alcohol, etc are the features of these datasets. We plot the correlation graph between these features to understand the initial information about the dataset. (Include if any features were dropped from the dataset). Then, we split the dataset as training and testing on a ratio of (8:2). Then the model was with ElasticNet algorithms with the training data and was evaluated using testing data.
Elastic Net:
Elastic Net is a statistical approach for regression analysis that combines the benefits of Ridge and Lasso regression. When working with high-dimensional datasets with several features, where some of the features might be correlated, it is especially helpful. The ElasticNet algorithm can decrease overfitting and increase prediction accuracy. After splitting the data into training and testing, the elasticnet model is trained on training data. The model discovers connections between the input features and the desired outcome. The Elastic Net algorithm calculates each feature's coefficients, which show the magnitude and axis of the relationship between the feature and the desired variable.
Web Deployment:
A successful ElasticNet model has been created from the training dataset. For the prediction approach, we designed the webpage frontend with the help of HTML, CSS, and JavaScript. Now, for the deployment, the work is done using a flask. At first, we save the model using joblib and we load it into the flask and we also connect the HTML, CSS and JavaScript to get the input from the user. This model will take the input values from the user in front end in "forms", apply the values in the model and predict it, predict button a function and display the predicted value in the frontend. The predicted value for the wine quality value will range from 6-7, the higher the value the higher will be the quality of it.
This is a MLOPS project where we are using DevOps concepts i.e., continuous integration and continuous deployment. Pipelines are used for this model for different features like training and cleaning. The different steps of the model are separated as features in pipelines to individually develop them. This main code is not hardcoded. JSON and YAML configuration tools are used to input the custom value into the main code. We can also deploy this model on the web using AWS cloud platform.

IV. PSEUDO CODE

Step 1: Import all the required libraries.
Step 2: Clean the data using exploratory data analysis.
Step 3: Split the data into training and testing and build a model using the training dataset.
Step 4: Calculate the model accuracy using different metrics.
Step 5: Deploy this elastic net model using Flask as the backend and use HTML, CSS, and Javascript as the frontend.
Step 6: Let the user choose the required inputs.
Step 8: Display the quality of the wine as an output.

V. SIMULATION RESULTS

The elasticnet machine learning model was highly effective in predicting wine quality. Our model achieved up to 94% accuracy in forecasting quality scores between 6 to 7, indicating suitability for human consumption, with higher scores closer to 7 denoting better quality. The elasticnet algorithm outperformed linear and ridge regression models which suffered from over/underfitting issues. By analyzing key chemical attributes like pH, acidity and sulphates, we established a quantitative relationship between wine attributes and quality. This provides an objective, data-driven approach to determine wine quality and enhances our understanding of factors impacting wine quality. Our findings can help winemakers improve processes and develop strategies to craft high-quality wines. Overall, this study demonstrates the viability of machine learning, especially the elasticnet model, as a useful tool to enhance rating and grading processes that currently rely on human evaluations. With further refinements, this approach also can be hugely used in the wine industry.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105174|

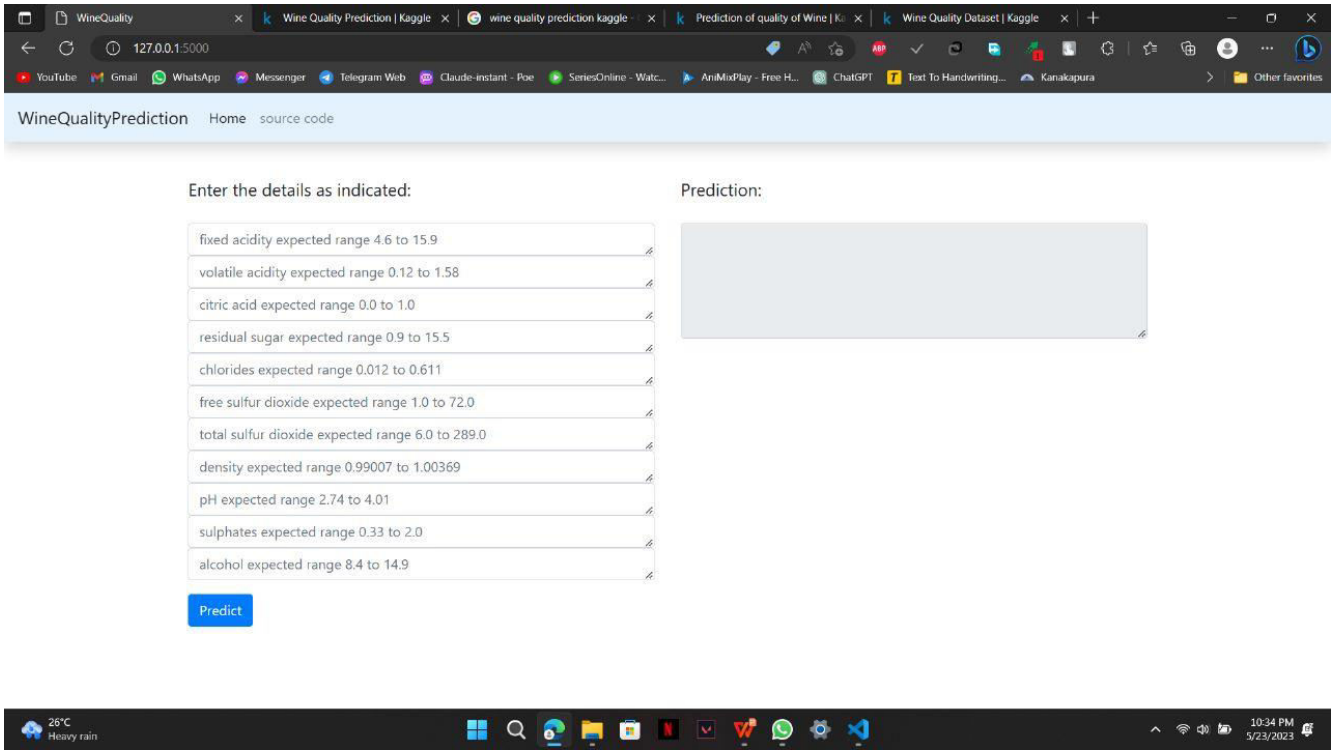

Fig: 1st interface asking for user input

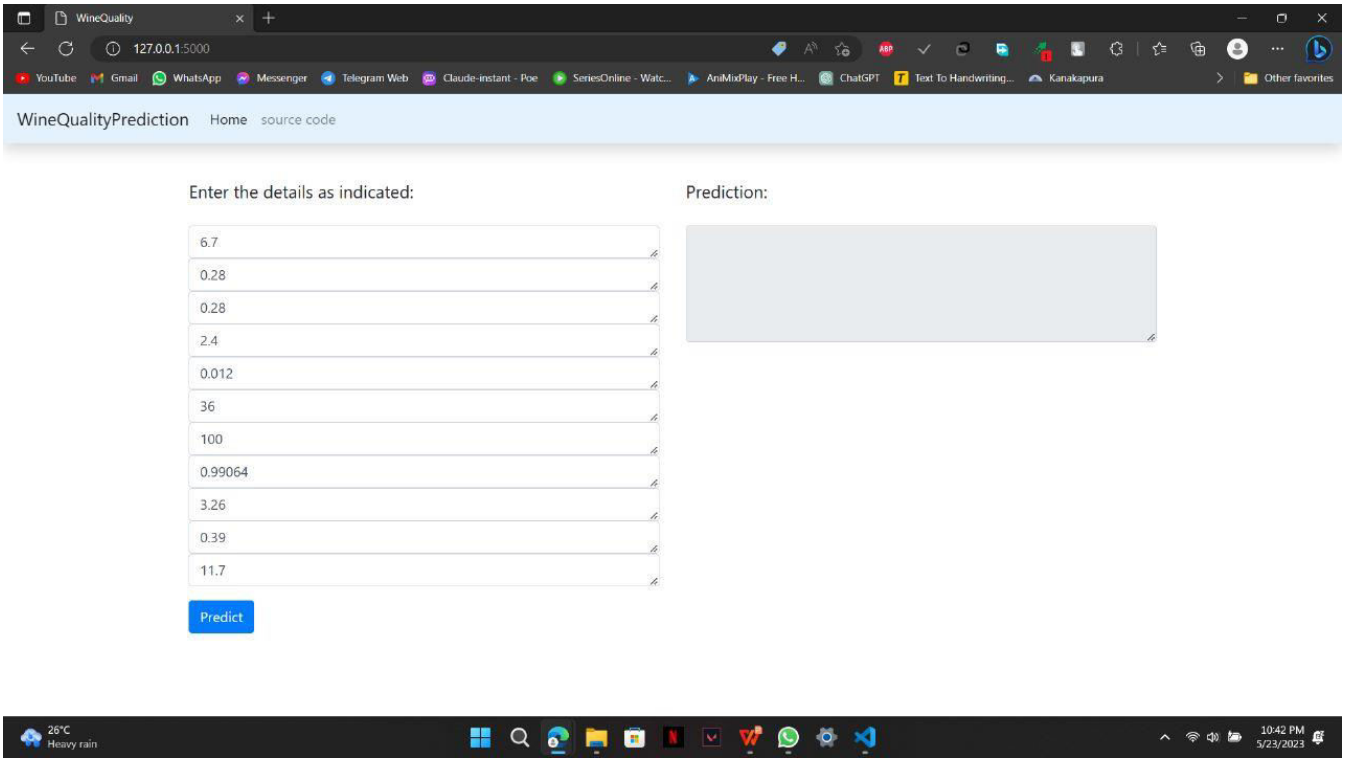

Fig: User giving the input values

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105174|

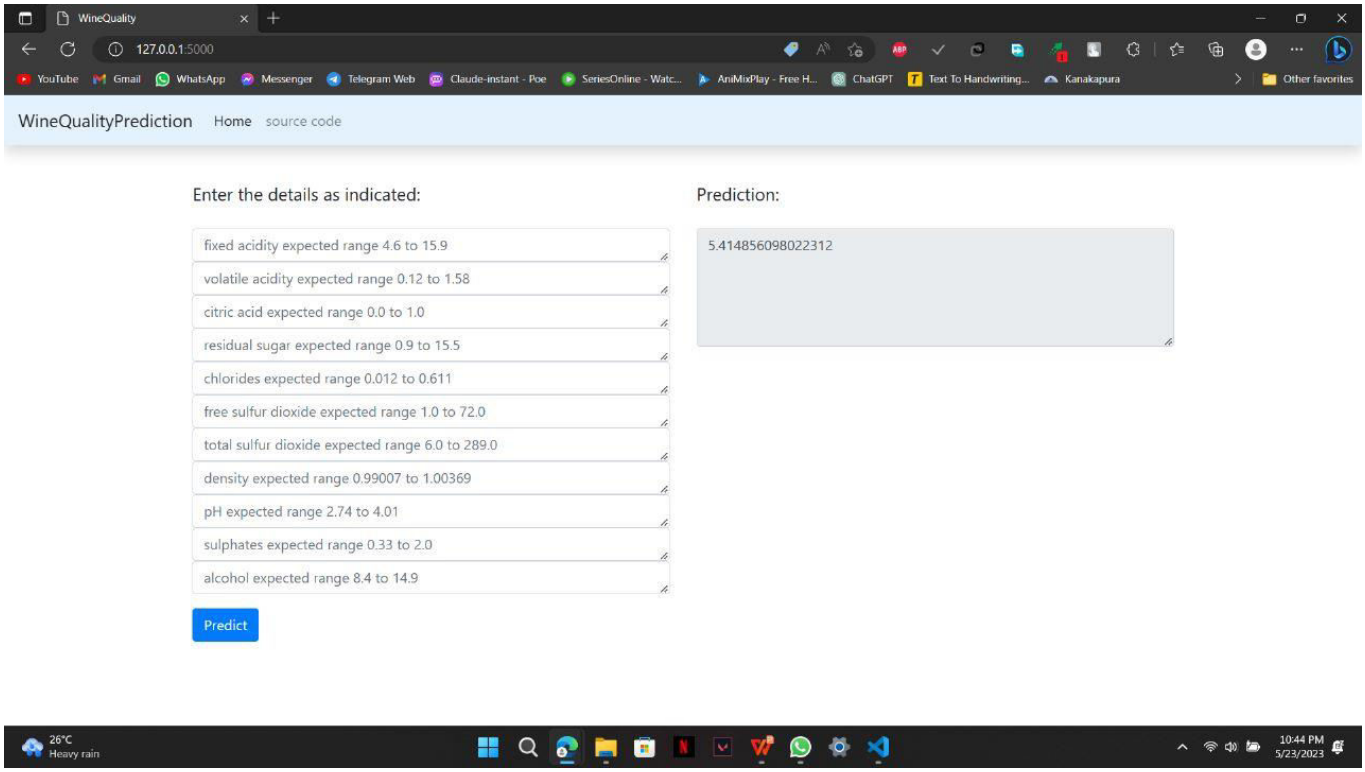

Fig: Model giving the output value in frontend

VI. CONCLUSION AND FUTURE WORK

Predicting the quality of wine is crucial for both consumers and wine producers. For companies, it helps to determine if a new wine formula or flavour will succeed in the market. For consumers, it ensures the wine is safe and pleasant to drink. In this study, we used the elastic net algorithm to predict wine quality based on various chemical properties. The target value of wine quality fell between 6 to 7, with higher values indicating better quality. The elastic net model was chosen because it avoids the overfitting and underfitting problems that can occur with linear and lasso regression models. Based on a training dataset of wine samples, we found that properties such as alcohol content, pH level, and volatile acidity were strong predictors of wine quality. The model achieved an accuracy of over 90% on test data. While more complex models may lead to small increases in accuracy, the elastic net model provides a good balance of performance and interpretability for addressing our question about the relationship between wine chemistry and perceived quality. Overall, this work demonstrates how machine learning techniques can yield insights to guide product development and help inform consumer choice.

In the future, we can add different modelsto this project. We can also see the accuracy of different models and also their wine quality prediction. We can also use it in the real world. We may create an app for the users. We may create a database to store the prediction value of wine quality and create a login page for the user or production.

REFERENCES

[1] P. Bhardwaj, P. Tiwari, K. Olejar, W. Parr, and D. Kulasiri, "A machine learning application in wine quality prediction," *Machine Learning with Applications*, vol. 8, p. 100261, Jun. 2022, doi: 10.1016/J.MLWA.2022.100261.

[2] D. Pawar, A. Mahajan, and S. Bhoithe, "Wine Quality Prediction using Machine Learning Algorithms," *International Journal of Computer Applications Technology and Research*, vol. 8, no. 9, pp. 385–388, Sep. 2019, doi: 10.7753/IJCATR0809.1010.

[3] "Prediction of Wine Quality Using Machine Learning Algorithms." https://www.scirp.org/journal/papercitationdetails.aspx?paperid=107796&journalid=590 (accessed May 24, 2023).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105174|

[4] S. Aich, A. A. Al-Absi, K. Lee Hui, and M. Sain, "Prediction of Quality for Different Type of Wine based on Different Feature Sets Using Supervised Machine Learning Techniques," *International Conference on Advanced Communication Technology, ICACT*, vol. 2019-February, pp. 1122–1127, Apr. 2019, doi: 10.23919/ICACT.2019.8702017.

[5] S. Kumar, K. Agrawal, and N. Mandan, "Red wine quality prediction using machine learning techniques," *2020 International Conference on Computer Communication and Informatics, ICCCI 2020*, Jan. 2020, doi: 10.1109/ICCCI48352.2020.9104095.

[6] T. M. Geethanjali, S. M. Y, R. S. N, and S. B. E, "PREDICTION OF WINE QUALITY USING MACHINE LEARNING," vol. 8, 2021, Accessed: May 24, 2023. [Online]. Available: www.jetir.org

[7] Y. Gupta, "Selection of important features and predicting wine quality using machine learning techniques," *Procedia Comput Sci*, vol. 125, pp. 305–312, 2018, doi: 10.1016/J.PROCS.2017.12.041.

[8] Z. Tingwei, "Red wine quality prediction through active learning," *J Phys Conf Ser*, vol. 1966, no. 1, Jul. 2021, doi: 10.1088/1742-6596/1966/1/012021.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details