Wine Quality Prediction Using IBM Watson Machine Learning

1. INTRODUCTION

OVERVIEW

Wine is the most commonly used beverage globally, and its values are considered important in society. Wine is an alcoholic drink that is made up of fermented grapes. Quality of wine is important for its consumers, mainly for producers in the present competitive market to raise the revenue. Wine quality refers to the factors that go into producing a wine, as well as the indicators or characteristics that tell you if the wine is of high quality. Historically, wine quality used to be determined by testing at the end of the production.

If you have come across wine then you will notice that wine has also their type, they are red and white wine. According to experts, wine is differentiated according to its smell, flavour, and colour, but we are not wine experts to say that wine is good or bad. Every person has their own opinion about the tastes, so identifying a quality based on a person's taste is challenging. Judging the quality of wine manually is a really tough task, even the professional wine tasters have the accuracy of 71%.

PURPOSE

In this project, we present a wine quality prediction technique that utilizes historical data to train simple machine learning models which are more accurate and can help us know the quality of wine. The models can be run on much less resource intensive environments. From this the best model is selected and saved in pkl format. We will be doing flask integration and IBM deployment.

2.LITERATURE SURVEY

EXISTING PROBLEM

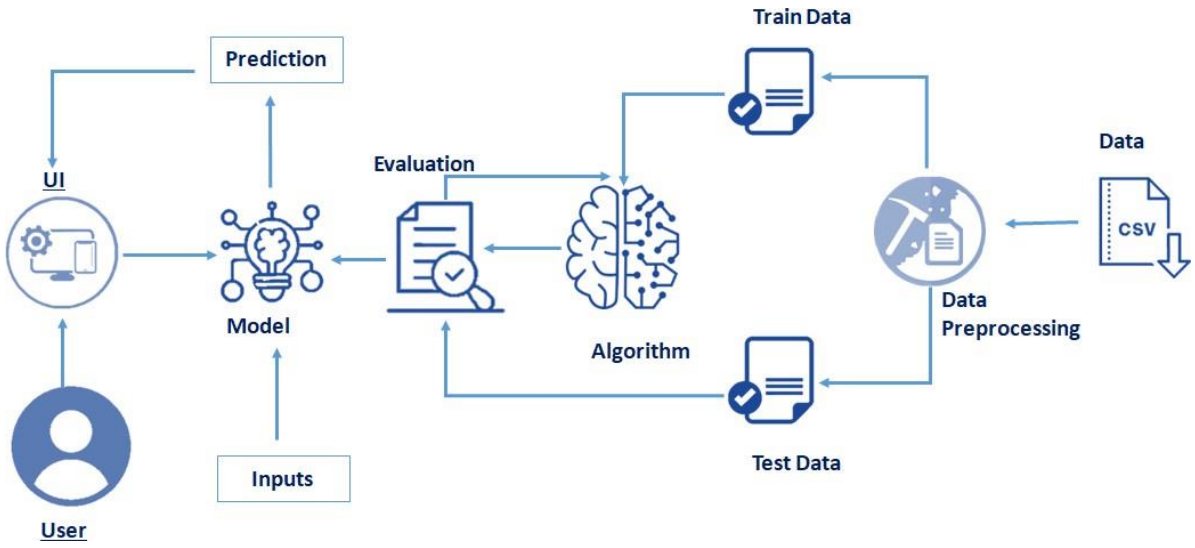
Ensuring the quality of wine whether you are testing for contaminants or developing a new product, we are here to help. Our knowledge, innovative products, and range of solutions allow our customers to maintain their focus where it should be – delivering high-quality wine that consumers expec

PROPOSED SYSTEM

Quality of wine is important for its consumers, mainly for producers in the present competitive market to raise the revenue. Wine quality refers to the factors that go into producing a wine, as well as the indicators or characteristics that tell you if the wine is of high quality. Historically, wine quality used to be determined by testing at the end of the production. we present a wine quality prediction technique that utilizes historical data to train simple machine learning models which are more accurate and can help us know the quality of wine. The models can be run on much less resource intensive environments

3. THEORETICAL ANALYSIS

BLOCK DIAGRAM



HARDWARE AND SOFTWARE DESIGNING

Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. It was created by Guido van Rossum , and first released on February 20, 1991. Its high-level built in data structures, combined with dynamic typing and dynamic binding , make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Anaconda Navigator

Anaconda Navigator is a free and open-source distribution of the Python and R programming languages for data science and machine learning related applications. It can be installed on Windows, Linux, and macOS.Conda is an open-source, crossplatform, package management system. Anaconda comes with so very nice tools like JupyterLab, Jupyter Notebook, QtConsole, Spyder, Glueviz, Orange, Rstudio, Visual Studio Code. For this project, we will be using Jupyter notebook and Spyder.

Jupyter Notebook

The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained bythe people at Project Jupyter. Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

Spyder

Spyder, the Scientific Python Development Environment, is a free integrated development environment (IDE) that is included with Anaconda. It includes editing, interactive testing, debugging, and introspection features. Initially created and developed by Pierre Raybaut in 2009, since 2012 Spyder has been maintained and continuously improved by a team of scientific Python developers and the community.Spyder is extensible with first-party and third party plugins includes support for interactive tools for data inspection and embeds Pythonspecific code. Spyder is also pre-installed in Anaconda Navigator, which is included in Anaconda.

Flask

Webframework used for building. It is a web application framework written in python which will be running in local browser with a user interface. In this application, whenever the user interacts with UI and selects emoji, it will suggest the best and top movies of that genre to the user.

Hardware Requirements:

o Operating system: window 7 and above with 64bit

o Processor Type -Intel Core i3-3220

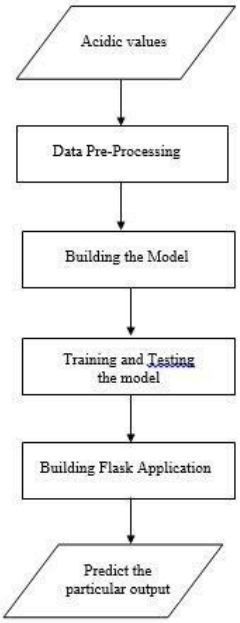
o RAM: 4Gb and above

o Hard disk: min 100GB

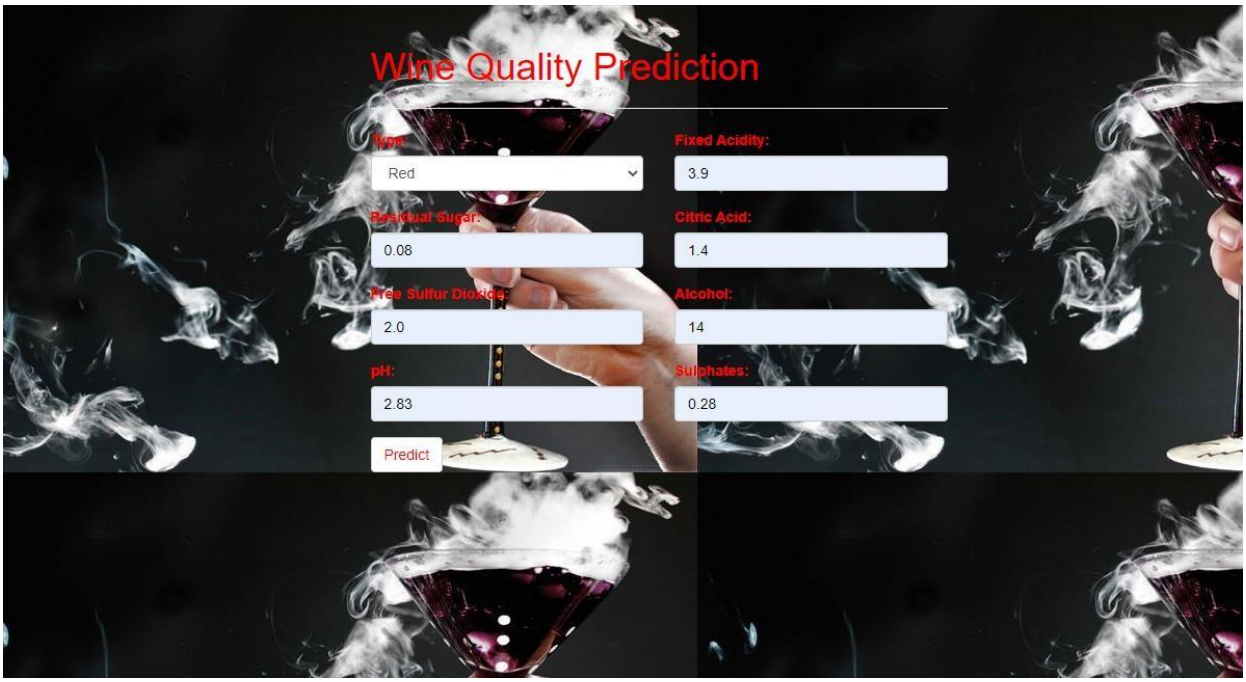
4. EXPERIMENTAL INVESTIGATION

Understanding the demands of wine safety testing can be a complex task for the laboratory with the numerous analytes and residues to monitor.Our separation and detection technologies, combined with experienced applications competence and our best suited chemistries, provide ideal solutions for the analysis of wine. Winemakers know, many of the quality improvements to wine each season, happen in the lab. Regular checks during production and being able to make quick decisions based on the results is key in achieving this quality – therefore, the accuracy and reliability of the data used to make those decisions is critica.

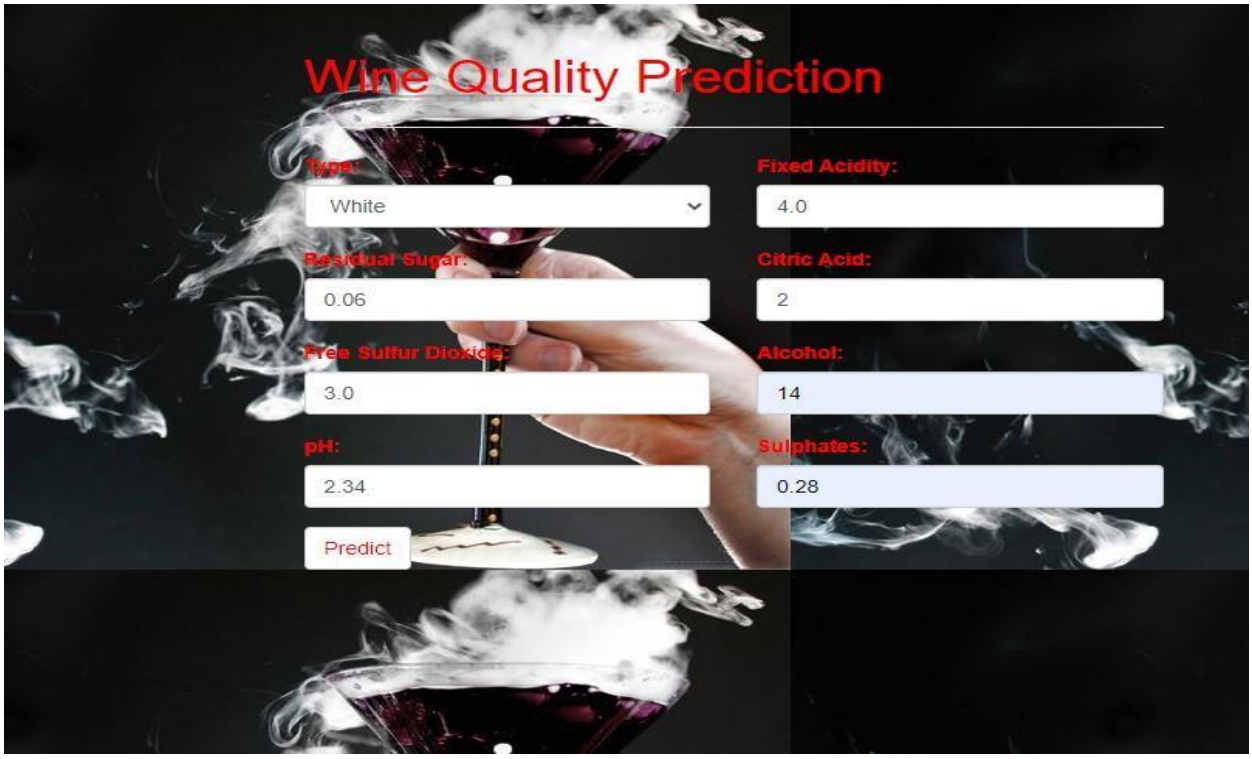
5. FLOWCHART



6. RESULT


Wine Quality Prediction
Type:
Red
Fixed Acidity:
3.9
Residual Sugar:
0.08
Citric Acid:
1.4
Free Sulfur Dioxide:
2.0
Alcohol:
14
pH:
2.83
Sulphates:
0.28
Predict


Wine Quality Prediction
Wine Quality is predicted to be :- Good


Wine Quality Prediction
Type:
White
Fixed Acidity:
4.0
Residual Sugar:
0.06
Citric Acid:
2
Free Sulfur Dioxide:
3.0
Alcohol:
14
pH:
2.34
Sulphates:
0.28
Predict


Wine Quality Prediction
Wine Quality is predicted to be :- Good

7.ADVANTAGES

- Easy to use
- Time Efficient
- Cost Efficient

8. APPLICATION

Work on several numbers of data: The number of choices for anything on internet is veryhigh and it's tedious to refine most wanted data by self while searching. The scope of this proposal system includes working within numerous data, with case.
Saving of time: Many people have problem selecting the alternative item of movie due to lack of time and due to search issues. Also movie recommendations from friends can be time consuming. The system helps in saving lots of time.

9 .CONCLUSION

First, we used oversampling to balance the dataset in the data preprocessing stage to optimize the performance of the model. Then we look for features that can provide better prediction results. For this, we used Pearson coefficient correlation matrices and ranked the features according to the high correlation among the features. After applying the sampling datasets which is balancing dataset the performance of the model is improved. In general, removing irrelevant features of the datasets improved the performance of the classification model. To conclude that the minority classes of a dataset will not get a good representation on a classifier and representation for each class can be solved by oversampling and undersampling to balance the representation classes over datasets. Therefore, in the classification algorithms by selecting the appropriate features and balancing the data can improve the performance of the model.

10. FUTURE SCOPE

In this project, we present a wine quality prediction technique that utilizes historical data to train simple machine learning models which are more accurate and can help us know the quality of wine. The models can be run on much less resource intensive environments. From this the best model is selected and saved in pkl format. We will be doing flask integration and IBM deployment.

11. APPENDIX

https://github.com/smartinternz02/SI-GuidedProject-319224-1664616759