Precise Coffee Quality Prediction

Milestone 1: Project Initialization and Planning Phase

The project initialization and planning phase for the coffee quality prediction project involves defining the objective, which is to predict the quality of coffee beans using machine learning. The scope includes data collection, analysis, and model deployment through a web application. Key steps include gathering comprehensive data on attributes such as variety, processing method, aroma, flavour, and defects, followed by thorough data cleaning to address any missing values or inconsistencies. Success is measured by the accurate prediction of coffee quality, leading to a functional, user-friendly web application for real-time assessments.

Activity 1: Define Problem Statement

The problem statement for the coffee quality prediction project is to develop a machine learning model capable of accurately classifying the quality of coffee beans based on various attributes such as aroma, flavour, aftertaste, acidity, body, balance, and defects. The goal is to create a reliable and precise tool that can assist coffee producers and quality control experts in identifying high-quality coffee beans and ensuring consistency in coffee production, ultimately enhancing consumer satisfaction and market value.

Coffee Quality Problem Statement Report: Click Here

Activity 2: Project Proposal (Proposed Solution)

The proposed solution for the coffee quality prediction project involves developing a robust machine learning model using a random forest algorithm to accurately classify coffee bean quality based on key attributes such as aroma, flavour, aftertaste, and defects. The project will encompass data collection, preparation, and analysis, followed by model training and validation. The final model will be deployed through a user-friendly web application, allowing real-time predictions and aiding coffee producers in maintaining high-quality standards and improving their product offerings.

Coffee Quality Project Proposal Report: Click Here

Activity 3: Initial Project Planning

Initial project planning for the coffee quality prediction project involves outlining key milestones and deliverables, including data collection, data preparation, exploratory data analysis, model building, and deployment. The plan will allocate resources and

responsibilities among the four team members, set timelines for each phase, and establish evaluation criteria for model performance.

Coffee Quality Project Planning Report: Click Here

Milestone 2: Data Collection and Preprocessing Phase

The data collection and preprocessing phase involves gathering comprehensive coffee quality data, including attributes like aroma, flavour, and defects, from reliable sources. The collected data is then cleaned to handle missing values and inconsistencies, ensuring it is ready for analysis and model building. This phase sets the foundation for accurate and reliable coffee quality predictions.

Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report

The data collection plan includes identifying raw data sources such as coffee farms, quality control centres, and industry databases to gather comprehensive data on attributes like aroma, flavour, and defects. A data quality report will be generated to assess the completeness, accuracy, and consistency of the collected data, ensuring it meets the standards required for reliable analysis and modelling.

Coffee Quality Data Collection Report: Click Here

Activity 2: Data Quality Report

The data quality report assesses the collected coffee quality data for completeness, accuracy, and consistency. It identifies and addresses missing values, outliers, and inconsistencies to ensure the dataset is reliable and ready for analysis and modelling, thereby ensuring the integrity of the subsequent machine learning predictions.

Coffee Quality Data Quality Report: Click Here

Activity 3: Data Exploration and Preprocessing

Data exploration and preprocessing involve analysing the collected coffee quality data to understand its distribution and relationships between attributes. This phase includes cleaning the data by handling missing values and outliers, normalizing or transforming variables as needed, and ensuring the dataset is structured and ready for effective model building and analysis.

Coffee Quality Data Exploration and Preprocessing Report Click Here

Milestone 3: Model Development Phase

The model development phase involves selecting and training a random forest algorithm using the pre-processed coffee quality data. This phase includes tuning hyperparameters, evaluating model performance with metrics such as accuracy and precision, and iteratively refining the model to ensure it accurately predicts coffee bean quality.

Activity 1: Feature Selection Report

The feature selection report identifies the most influential attributes from the coffee quality dataset that contribute significantly to the random forest model's predictive accuracy. It highlights key features such as aroma, flavour, and defects, based on their importance scores derived from the model, ensuring the inclusion of critical variables for robust coffee quality predictions.

Coffee Quality Feature Selection Report: Click Here

Activity 2: Model Selection Report

The model selection report evaluates various machine learning algorithms and concludes that the random forest model is optimal for predicting coffee bean quality based on attributes such as aroma, flavour, and defects. It outlines the reasons for selecting this model, including its ability to handle complex relationships in the data, robustness against overfitting, and high predictive accuracy demonstrated during testing and validation phases.

Coffee Quality Model Selection Report: Click Here

Activity 3: Initial Model Training Code, Model Validation and Evaluation Report

The initial model training code utilizes a random forest algorithm to train on preprocessed coffee quality data, incorporating attributes like aroma, flavour, and defects for predicting bean quality. The model validation and evaluation report subsequently assess its performance through metrics such as accuracy, precision, and recall, confirming its robustness and suitability for real-world deployment in predicting coffee bean quality effectively.

Coffee Quality Model Development Phase Template: Click Here

Milestone 4: Model Optimization and Tuning Phase

During the model optimization and tuning phase, the random forest algorithm parameters are adjusted to enhance performance and accuracy in predicting coffee bean quality. This involves fine-tuning hyperparameters such as tree depth and number of estimators, optimizing feature selection, and employing cross-validation techniques to ensure the model's reliability and effectiveness across different datasets.

Activity 1: Performance Metrics Comparison Report

The performance metrics comparison report evaluates and compares key metrics such as accuracy, precision, and recall across different models or variations of the random forest algorithm. It highlights the optimal configuration that achieves the highest predictive accuracy and reliability in determining coffee bean quality based on attributes like aroma, flavour, and defects

Activity 2: Final Model Selection Justification

The random forest model was selected based on its proven ability to effectively predict coffee bean quality by leveraging attributes such as aroma, flavour, and defects. It demonstrated superior performance through rigorous evaluation, ensuring reliable and precise predictions essential for enhancing coffee quality assessment and production standards.

Coffee Quality Model Optimization and Tuning Phase Report: Click Here

Milestone 5: Project Files Submission and Documentation

For project file submission in GitHub, kindly click the link and refer to the flow.
Click Here

For the documentation, kindly refer to the link: Click Here

Milestone 6: Project Demonstration

In the README file uploaded the project demonstration video link. Explained the project and demonstrated its execution during the presentation.