Model Development Phase Template

Date	8 July 2024
Team ID	740074
Project Title	Rhythmic Revenue: Unveiling The Future Of Music Sales With Machine Learning
Maximum Marks	6 Marks

Model Selection Report

In the forthcoming Model Selection Report, various models will be outlined, detailing their descriptions, hyperparameters, and performance metrics, including Accuracy or F1 Score. This comprehensive report will provide insights into the chosen models and their effectiveness.

Model Selection Report:

Model	Description	Hyperparameters	Performance Metric (e.g., Accuracy, F1 Score)
Decision tree	In the context of predicting music sales, a decision tree model helps in making predictions based on various features such as genre, artist popularity, release date, and historical sales data.For this project, the decision tree helps identify the factors that influence music sales the most and predicts future sales based on historical data.	-	Training accuracy 99% and test accuracy 94%
Random forest	Random Forest is an ensemble learning method that enhances the predictive accuracy and robustness of decision trees by constructing multiple trees during training and outputting the average prediction (for regression) or majority vote (for classification) of the	-	Training accuracy 99% and Testing accuracy 94%

	individual trees. In the context of predicting music sales, the Random Forest model leverages various features such as genre and historical sales data to provide accurate and reliable forecasts.		
Linear Regression	linear regression helps forecast future music sales by establishing a linear relationship between sales and various influencing factors such as marketing spend, social media engagement and historical sales trends. The simplicity and interpretability of linear regression make it a valuable tool for understanding how different features contribute to sales predictions and for making informed, data-driven decisions.	-	Training accuracy 18% and Testing accuracy 20%
KNN	the K-Nearest Neighbors (KNN) algorithm is employed to predict music sales based on the proximity of data points in the feature space. KNN is a simple, yet effective, instance-based learning method where the model makes predictions by identifying the 'k' closest data points (neighbors) to a given input and then taking a majority vote (for classification) or averaging the values (for regression). This approach is particularly useful for identifying patterns and trends in music sales data by leveraging historical sales and other relevant features.	-	Training accuracy 74% and Testing accuracy 60%
XGBoost	XGBoost (Extreme Gradient Boosting) is a powerful ensemble learning method used for both classification and regression tasks. It builds multiple decision trees in a sequential manner, where each new tree corrects the errors of the previous ones. This approach results in a highly accurate and robust model. For the Rhythmic Revenue	-	Training accuracy 99% and Testing accuracy 94%

	project, XGBoost efficiently handles large datasets and captures complex patterns in music sales data, leading to precise predictions of future sales. Its ability to regularize helps in preventing overfitting, making it a reliable choice for our predictive analysis.		