

Data Collection and Preprocessing Phase

Date	15 JULY 2024
Team ID	SWTID1720151909
Project Title	PANIC DISORDER SETECTION
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	The data set consist data from 1906-2015 which will not be the correct estimation for further years	Moderate	<p>To address the issue of outdated data for future predictions, we can use advanced statistical and machine learning techniques. Here's a technical solution:</p> <ol style="list-style-type: none"> Data Preprocessing: <ul style="list-style-type: none"> Feature Engineering: Create new features that capture trends and seasonality. Data Normalization: Normalize the data to make it suitable for modeling. Model Selection: <ul style="list-style-type: none"> Use time series forecasting models like ARIMA, SARIMA, or Prophet, which are

			<p>specifically designed to handle time-dependent data and can capture trends and seasonality.</p> <ul style="list-style-type: none">○ Use regression-based models like Linear Regression, Random Forest, or XGBoost with temporal cross-validation to predict future values. <p>3. Training and Validation:</p> <ul style="list-style-type: none">○ Split the data into training and validation sets based on time. Ensure that the validation set consists of the most recent data.○ Use techniques like walk-forward validation to iteratively train and validate the model on multiple subsets of the data to ensure robustness. <p>4. Prediction:</p> <ul style="list-style-type: none">○ Generate future predictions using the trained model.○ Evaluate the predictions using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared to assess model performance. <p>5. Model Updating:</p> <ul style="list-style-type: none">○ Continuously update the model with new data as it becomes available to improve accuracy and relevance.
--	--	--	--