Report on Solving the Assignment: Building Statistical and Deep Learning Models for Google Play Store Dataset

Name:Muhammad Shaheer(20p-0480)

Instructor:Sir Shahzaib

This report provides a step-by-step guide on how to solve the assignment involving building both statistical and deep learning models for the Google Play Store dataset.

# Step 1: Data Preprocessing

The first step is to preprocess the dataset to ensure data quality and readiness for model training. This includes the following sub-steps:  
- Load the dataset using pandas and display its first few rows.  
- Handle missing values by using SimpleImputer to fill numerical columns with the median value.  
- Convert non-numeric columns to numeric where necessary and handle errors by coercing them to NaN.  
- Fill NaN values in the 'Genres' column with 'Unknown' to prevent issues during encoding.  
- Normalize numerical features using StandardScaler for standardization.  
- Remove any duplicate entries to maintain data integrity.

# Step 2: Building the Statistical Model

In this step, we build a statistical model using Logistic Regression:  
- Convert categorical labels ('Genres') to numerical values using the factorize method.  
- Split the dataset into training and test sets using train\_test\_split.  
- Train the Logistic Regression model and evaluate its performance on the test set.  
- Calculate and print the accuracy, precision, recall, and F1-score of the model.

# Step 3: Building the Deep Learning Model

This step involves building a deep learning model using Keras:  
- Prepare the data by encoding the target variable ('Genres') and splitting it into training and test sets.  
- Define the neural network architecture using the Sequential model, including Dense and Dropout layers.  
- Compile the model using the 'adam' optimizer and 'sparse\_categorical\_crossentropy' loss.  
- Train the model on the training data and validate it using a validation split.  
- Evaluate the model's performance on the test set and print the test accuracy.

# Step 4: Visualizing Model Performance

To better understand how the model performed, we create visualizations:  
- Plot the training and validation accuracy over epochs.  
- Plot the training and validation loss over epochs.  
- Use matplotlib to generate the plots for visual inspection.

# Step 5: Building the Recommendation System

The final step involves building a recommendation system:  
- Select 3 random applications from any genre.  
- Filter applications based on user input such as category, content rating, and whether they are free or paid.  
- Ensure at least 7 recommendations are available by falling back to filtering by content rating if necessary.  
- Choose 7 applications based on user preferences and ensure there are no duplicates.  
- Return the final list of recommended applications.

# Conclusion

This report outlined the process for solving the assignment, from data preprocessing to building and evaluating models, to developing a recommendation system. By following these steps, we ensure that the models are trained on high-quality data and produce meaningful results for analyzing the Google Play Store dataset.