**Customer Engagement Analysis Report**

**Overview**

The project is centred around forecasting customer engagement with offers in the Retail Industry, explicitly focusing on Starbucks. The main objective is to predict how customers would respond to different offers. This is crucial for any retail business as it directly impacts customer retention and sales.

The project uses a Random Forest model, a popular machine learning algorithm, to make these predictions. Random Forest is an ensemble learning method that operates by constructing multiple decision trees at training time and outputting the class, which is the mode of the classes of the individual trees. It is particularly well-suited for this task due to its ability to handle large datasets with high dimensionality and its robustness against overfitting.

The project involves extensive data analysis of customer profiles, transaction records, and offer details. The customer profiles provide demographic information about the customers, the transaction records contain information about the customers’ purchase history, and the offer details include information about the different offers sent to the customers.

The analysis aims to uncover patterns and relationships in the data that can help predict customer behaviour. For example, it might reveal that customers of a specific age group or income level are more likely to respond to offers. These insights can then be used to tailor the offers to different customer segments, thereby increasing the effectiveness of the marketing strategy.

**Data**

Starbucks provides the Starbucks dataset to Udacity scholars for creating innovative projects. The dataset has three separate files: portfolio.json, profile.json, and transcript.json.

**Portfolio.json**: This file contains data variables such as reward (int), difficulty (int), and duration (int) as continuous values. It also includes channels (list of strings), offer\_type (str), and id (str) as categorical data. This data provides information about the different offers sent to the customers.

**Profile.json**: This file contains data variables such as age (int), became\_member\_on (int), and income (float) as continuous values. It also includes gender (str) and id (str) as categorical data. This data provides demographic information about the customers.

**Transcript.json**: This file contains data variables such as time (int) as a continuous value. It also includes person (str), event (str), and value (dict of strings) as categorical data. This data provides information about the customers’ transaction history.

**Methodology**

This project's methodology uses a **Random Forest model** to predict **customer engagement** with offers.

Here are the key steps involved:

**Data Preprocessing:** The first step in the methodology is data preprocessing. This involves cleaning the data, handling missing values, and transforming the data into a suitable format for the model. The data is processed to create a binary indication of whether a customer responded to an offer, with 1 indicating that the customer viewed and completed the offer and 0 indicating that the offer was considered but not used.

**Feature Engineering:** The next step is feature engineering. This involves creating new features from the existing data that can help improve the model's performance. For example, features could be made to represent the total amount spent by a customer, the average amount spent per transaction, the total number of transactions, etc.

**Model Training:** Once the data is pre-processed and the features are created, the next step is to train the Random Forest model. The model is trained on the training data, which includes the features and the target variable (whether the customer responded to the offer).

**Hyper Parameter Tuning:** The Random Forest classifier was then tuned with the following parameters: **Max Features, Depth of the tree, number of estimators and Bootstrapping.**

**Model Evaluation:** After the model is trained, it is evaluated on the test data to assess its performance. The model's performance is evaluated using appropriate metrics, such as accuracy, precision, recall, and F1 score.

**Prediction:** Once the model is trained and evaluated, it can be used to make predictions on new data. The projections can provide insights into how customers will likely respond to different offers.

This methodology allows for a systematic approach to predicting customer engagement with offers. It ensures that the model is trained on high-quality data and its performance is thoroughly evaluated.

**Results**

The project's results are evaluated based on the performance of the Random Forest model on the test dataset. The model outputs the predicted response of the customers to the offers, and these predictions are compared with the actual responses to evaluate the model's performance.

The key metric to evaluate the model's performance is accuracy, the proportion of correct predictions out of the total predictions made. The model achieved an overall accuracy of **92.3%** in testing, a strong result for a task as complex as predicting customer engagement with offers.

In addition to accuracy, other metrics such as precision, recall, and F1 score could also be used to evaluate the model's performance. Precision measures the proportion of accurate positive predictions out of all optimistic predictions, recall measures the proportion of accurate positive predictions out of all actual positives, and the F1 score is the harmonic mean of precision and recall. These metrics provide a more comprehensive view of the model’s performance.

The results also include a feature importance analysis, identifying the most influential features in predicting customer engagement. This analysis can provide valuable insights into the factors that drive customer engagement and can inform future strategies to improve customer engagement.

**Conclusion**

The project's conclusion encapsulates the work's significance and impact. The project demonstrates the practical application of machine learning models, specifically the Random Forest model, in predicting customer behaviour in the retail industry.

Despite the challenges posed by the complexity of the task and the high dimensionality of the data, the model provides reliable and accurate predictions. This is primarily due to the robustness of the Random Forest model against overfitting and its ability to handle large datasets with high dimensionality.

The project also involves extensive data analysis of customer profiles, transaction records, and offer details. The insights gained from this analysis can help businesses understand their customers better and tailor their offers to different customer segments, thereby increasing the effectiveness of their marketing strategies.

In conclusion, the project successfully aims to develop a robust and reliable system for predicting customer engagement with offers. It showcases the potential of machine learning models in handling complex tasks and sets the stage for further improvements and innovations in this field.