PART 1: PROBLEM DEFINITION AND DATASET PREPARATION

Customer retention is increasingly recognised as a critical challenge across different industries. It is not a secret that several factors drive customer turnover, including competitive offerings at lower prices, negative influences from word-of-mouth or social media, superior customer service from competitors, among others. Studies highlight that it is more cost-effective to retain an existing customer than to acquire a new one, mainly due to the marketing expenses associated with attracting new customers (Arslan, 2020). Accordingly, in today's competitive market, securing a loyal customer base is essential. Customer churn usually occurs gradually, not suddenly, indicating the possibility of taking a proactive approach by analysing historical purchase patterns to foresee churn. According to Dahiya and Bhatia (2015), customer churn refers to customers who have a high probability to stop transacting with a company. However, thanks to the detailed recording and storage of transaction data, it is feasible to gain an in-depth understanding of customer behaviours and preferences, facilitating the development of effective retention strategies.

Within the scope of this report, let's assume a critical challenge facing Tesco's sales and marketing department, particularly their method of managing customer retention. Currently, their strategy is to engage with customers only after they have ceased their shopping activities, with efforts being sporadic and lacking a cohesive strategy. This reactive method has proven ineffective, often failing to reclaim customers once they have departed. As a Tesco Business Analyst, I am tasked with developing a system to detect early signs of disengagement among our most loyal customers. This initiative aims to pre-emptively counter potential customer attrition. Initially, my analysis will focus on Tesco’s transactional data from December 2009 to December 2011 to identify patterns in shopping frequency, expenditure and the duration of the customer's relationship with Tesco (tenure). The main goal is to identify loyal customers who reduce their shopping basket's value to £0 over two consecutive months, signalling potential churn. For example, a customer who made purchases in April 2010 but did not shop in both May and June 2010 would be flagged as a potential churn risk. Subsequently, I will create a deep learning model using a Recurrent Neural Network (RNN) architecture enhanced with Long Short-Term Memory (LSTM) to forecast customer activities, like purchases patterns, over upcoming time intervals. These predictions, indicating anticipated customer activity levels, will be normalised to a 0-1 range so they can be interpreted as probabilities. Part 2 will provide more information regarding model development and application.

Before proceeding with the development of the model, the dataset underwent several key preprocessing steps to ensure its readiness for modelling. Firstly, I merged the two sheets "Year 2009-2010" and "Year 2010-2011" to create a comprehensive, longitudinal view of customer purchasing behaviour. Following this, I inspected all columns for missing values. Special attention was given to the 'Customer ID' field, where any missing IDs led to the removal of corresponding rows, as tracking customer behavior without a unique identifier is not feasible. Afterwards, I identified and removed rows featuring negative or zero quantities, which typically represent returns or other non-purchases and do not contribute to understanding genuine purchasing behaviour. Subsequently, I calculated the average number of months per year in which each customer made at least one purchase. This average was then included in our dataset as a 'Purchase Frequency Score'. This new metric serves as a criterion to distinguish our most loyal customers from those who shop sporadically. After doing this, I systematically evaluated customer activity and churn risk over a series of months, focusing on understanding patterns of purchasing behaviour.

Moving forward, I created a new Dataframe to classify each customer's monthly purchase activity. Initially, each month for each customer is labelled as either active (if there's a purchase) or inactive (if there isn't). The labels are refined to indicate churn risk. If a customer has no purchases in a given month and also no purchases in the following two months, they're labeled as high churn risk (2). Other conditions result in a label of low churn risk (1) or active (0). For the last two months of data, I used a different approach since it's not possible to look ahead two months. These months are labelled as active (0) or inactive (1) based on that month's purchase activity alone. Another important pre-processing step was identifying our most loyal customers based on their purchase frequency score. We selected customers in the top quartile (0.75), which includes those who, on average, made at least one purchase every six months per year.

PART 2: MODEL DEVELOPMENT AND APPLICATION

As mentioned earlier, I developed a deep learning model using an RNN architecture enhanced with LSTM units to leverage their ability to remember long-term dependencies, which is crucial for learning sequential patterns from customer data changing over time. Our input features, denoted as X, comprise the monthly purchase values for each customer. The target outputs, Y (Labels), consist of monthly churn risk labels for each customer, which the model aims to predict. Subsequently, I divided the dataset into training, validation and test sets to ensure robust model evaluation and to mitigate overfitting. The model was constructed using

the Keras Sequential API. The input layer is designed to accommodate sequences of 24 timesteps, corresponding to our time series data, with one feature per timestep. The model includes two LSTM layers, each with 20 neurons. Both layers are configured to return sequences, allowing the network to continue learning from the sequence in subsequent layers. This setup ensures that each LSTM layer contributes to a deeper understanding of the sequential patterns in the data. The time distributed dense layer serves as the output layer, encapsulating a single-unit dense layer that is applied independently to each timestep of the LSTM's output. This configuration allows for a prediction at every timestep, aligning the model's output sequence directly with the length of the input sequence. By doing so, it ensures that the model delivers a continuous sequence of predictions, each reflecting the temporal dynamics learned up to that point in the sequence.

The model employs the Adam optimiser with a learning rate of 0.001 and uses Mean Squared Error (MSE) as the loss function, which measures the average squared differences between predicted and actual values. It also tracks Mean Absolute Error (MAE) to assess the average magnitude of errors without considering their direction. Training was conducted over up to 1,000 epochs with a batch size of 32, but early stopping was triggered at 179 epochs when the validation loss ceased to improve, ensuring the model didn't overfit by reverting to the best weights observed. Upon evaluation on a separate test dataset, the model achieved a test MSE of 0.063 and an MAE of 0.138, indicating a reasonable fit to the data. The small difference (approximately 0.002) between training and validation MAE suggests low variance and confirms that the model generalises well without significant overfitting. The Root Mean Squared Error (RMSE) on the test set was approximately 0.25, providing a further measure of prediction error. Overall, the model is deemed appropriate for the task, though there is potential for enhancement. Possibilities for improvement include expanding the dataset with additional historical data to better capture underlying trends or refining the model architecture, despite previous attempts at optimisation.

Lastly, I made predictions using the full dataset that was employed to construct our X, normalising these predictions between 0 and 1, and specifically focused on the predictions for the last month. This focus is due to how LSTM models initialise their states with zeros and gradually build up their internal state as they process the sequence. Early predictions may be less reliable because the model hasn't fully developed its internal state. By the last month, the LSTM has processed all previous data, creating a rich internal state that more accurately captures underlying patterns, making these later predictions more trustworthy.

PART 3: ETHICAL CONSIDERATIONS

Fairness: the model's predictions depend heavily on the data it was trained on. In our case, dataset is predominantly composed of transactions from UK customers. This data representation could inadvertently lead the model to develop predictions that are more accurate for UK-based shopping patterns, preferences and behaviors, which may not necessarily align with those of customers from other countries. Consider a scenario where the company has a growing customer base in other parts of Europe, like Germany and France. Customers in these countries might have different shopping frequencies, basket sizes or preferences due to cultural or economic differences. However, since the model is primarily trained on data from UK customers, it might not capture these nuances effectively. Consequently, the model may inaccurately predict churn for non-UK customers because it underestimates or misinterprets their engagement patterns and the company might deploy retention strategies that are either unnecessary or ineffective for these customers.

Accountability: it should be possible to explain the model's decisions to stakeholders, including customers. Since the model predicts churn risk as a continuous probability rather than a binary outcome, it's important to clearly communicate what these probabilities represent to both internal stakeholders and customers. For example, a churn risk score of 0.8 does not necessarily mean a customer is definitely leaving but indicates a high likelihood. The factors contributing to this score should be transparent. Moreover, to manage predictive uncertainty, it is important to establish confidence thresholds for churn predictions. These thresholds will determine the actions triggered when certain levels of risk are identified. Accordingly, there should be mechanisms to handle errors in predictions. Essentially, customers impacted by such errors should have a recourse to contest or correct decisions made based on the model's output.

Transparency: customers must be informed about how their data is being used to model their behaviour, especially in contexts that may affect their perceived relationship with the company. This is crucial for maintaining trust and consent in the use of their personal information. The purpose and function of the model, like how it works, what data it uses, and how these contribute to decisions, should be transparently communicated to all stakeholders.

PART 4: GENERATIVE AI IN BUSINESS

A creative solution to early detection and prevention of disengagement among loyal customers involves using generative AI to create automated, personalised retention offers. This system designs offers tailored to individual customers based on their profiles, shopping behaviours and patterns indicating potential churn. This system would leverage the transactional and

behavioural data that has already been collected and analysed to predict churn. This is a new component which uses generative adversarial networks or variational autoencoders trained on successful past offers (from historical data) and customer responses to generate creative and personalised offers. It monitors the acceptance and effectiveness of offers to refine and improve the generation algorithms.

The model provides several advantages, including increased customer retention by tailoring offers to address the specific factors leading to customer disengagement. It enhances customer satisfaction through personalised interactions that improve the overall customer experience and loyalty. Additionally, it aids in revenue preservation as retaining existing customers is often more cost-effective than acquiring new ones. Conversely, the main challenges associated with this model are that it must ensure compliance with data privacy and security regulations. Moreover, the integration of various AI technologies, like predictive analytics and generative models, introduces significant complexity. The generated offers need to balance creativity with operational feasibility to ensure they can be successfully implemented. Finally, the model requires the capability to swiftly adapt to changes in consumer behaviour and market conditions to remain effective.

REFERENCES

Arslan, I.K. (2020) "The importance of creating customer loyalty in achieving sustainable competitive advantage", *Eurasian Journal of Business and Management*, 8(1), pp. 11-20.

Dahiya, K. and Bhatia, S. (2015) "Customer churn analysis in telecom industry", *4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, pp. 1-6