MSc Data Analytics

Can Machine Learning Detect Psychological Biases in Consumer Purchase Behaviour?

A behavioural data approach to predictive modelling and interpretation

---

Dissertation

By

Davina Khimani (UP2089114)

Supervisor

Dr Ella Haig

September 2025

Abstract

Psychological biases tend to be implemented by marketing companies in e-commerce environments to influence consumer purchases. Machine learning (ML) has shown promise in predicting purchase behaviour, although, there is still some disconnect between computational prediction and behavioural theory. This dissertation examines if machine learning may detect the influence of psychological biases in predicting consumer behaviour, operationalising these biases as engineered features and interpreting them using explainable AI (SHAP).

Three datasets were utilised to predict urgency bias induced purchases (Amazon transactions, n=1.8 million), social proof (women's clothing reviews, n = 23k) and framing bias (synthetic e-commerce data, n = 100k). Engineered features capture behavioural cues, including holiday proximity reflecting a sense of urgency, review-based social proof scores and discount framing. Six machine learning models were used: Logistic Regression, SVM (linear models), Random Forest, XGBoost (ensemble models) and two neural networks; one simple and one enhanced.

Results indicated ML's ability to detect the influence of these biases but effectiveness varied. Framing bias produced near-perfect predictions across all models, although raw transactional variables outweighed the influence of engineered discount features. Purchases influenced by urgency bias had high accuracy (~92-95%), although recall scores were limited due to class imbalance, decreasing accurate predictions of urgent instances. Social proof bias performed the worst with accuracy plateauing to 83% across all models, due to noisy review data. SHAP feature importance revealed some biases aligned with behavioural theory, especially urgency features and some social proof features, though, framing proxies had significantly less influence compared to raw features.

The study emphasises the importance of data quality and feature design rather than model complexity in incorporating psychological theory into ML. It provides further growth and advancement in the possibility of bias-aware predictive modelling, while highlighting the need for enhanced feature engineering and data quality in future research to better expand this field of research.

Word Count: 12,975

List of Figures

List of Tables

Table of Contents

Chapter 1:
Introduction

1.1 **Overall Problem and Justification of Topic**

Every click, every product review, every discount in digital marketplaces carries more weight than it seems – small promotional cues can trigger powerful psychological biases that lead consumers to impulsively spend money and companies to strategically exploit these impulsive decision shortcuts. Consumer behaviour in e-commerce is heavily influenced by psychological biases, rather than consumers' rationally evaluating the product quality and pricing. Individuals deviate from rational choice models with behavioural economic theories such as Prospect Theory (Kahneman & Tversky, 1979) and Cialdini's Principle of Influence (Cialdini, 2001) highlighting how biases such as urgency, social proof and discount framing systematically affect decision-making. Digital marketplaces capitalise on these biases through persuasive techniques such as limited time offers, product ratings, advertising product value with discount and many more that influence millions of transactions daily.

Machine learning has made tremendous progress in predicting consumer purchases (Balyemah et al., 2024; Lin, 2025), although it has been critiqued for putting prediction ahead of interpretability. Features like 'discount percentages' or 'review counts' that are often used in models are rarely regarded as behavioural constructs – causing a gap between predictive modelling and behavioural theory.

The key issue this dissertation addresses is if machine learning can be used to detect the influence of psychological biases in consumer decisions. To assess the impact of behavioural cues, bias-aware feature engineering is required along with explainable AI techniques such as SHAP to measure feature importance. This may be technically challenging but an important task as it enhances the link between computational modelling and actual human-decision making.

This investigation is relevant given e-commerce environments rely significantly on persuasive marketing techniques to influence impulsive consumer purchases, yet not many computational models explicitly represent consumer biases. The novelty of this study lies in bias-aware feature engineering which directly encodes psychological constructs into machine learning pipelines, as well as the use of SHAP to meaningfully interpret predictions. The challenge remains that psychological biases cannot be directly observed in the data, they must be interpreted from proxies like purchase timings, review patterns or relative discounts. Additionally, it is difficult to balance model performance with interpretability, where some models such as neural networks are very accurate and perform optimally but it is difficult to explain given its complex structure.

By addressing these challenges, this dissertation presents a paradigm for detecting the influences of biases with machine learning, with implications for industrial practice and research.

1.2 Project Aims & Objectives

The aim of this dissertation is to investigate whether machine learning can be used to predict the influence of psychological biases in consumer purchase behaviour. So, this project will operationalise urgency, social proof and framing biases within e-commerce datasets.

Key Objectives:

1. To feature engineer existing dataset features to represent urgency, social proof and framing bias
2. To train and evaluate many machine learning models: Two traditional baseline models (Logistic Regression and SVM); two ensemble tree-based models (Random Forest and XGBoost) and two neural networks (Simple Neural Network and an Enhanced Neural Network).
3. To apply explainable AI tools (SHAP) to infer which features are the most influential to prediction outcomes.

4. To assess the feasibility of bias-aware predictive modelling by comparing performance across models and biases.

Achieving all these objectives will aim to provide both theoretical validation and predictive insight into how computational models can reflect underlying psychological processes.

1.3 Dataset Intro & Ethical Issues

This project utilised three datasets to model urgency, social proof and framing biases. For urgency bias, a large-scale Amazon transactions dataset was used (*n = 1.8 million*). A women's clothing review dataset (*n=23,486)* was used to model social proof, consisting of product ratings, recommendations and review helpfulness scores. Framing bias was captured using a synthetic e-commerce dataset (*n=100,000*) containing variables such as discount amount, units sold and revenue. Each dataset was pre-processed then feature engineered to represent its bias constructs: urgency labels near holidays, social proof scores that combine review quality and endorsement, and relative discount framing features.

Ethical considerations include:

- Data privacy – all datasets are publicly available but remained anonymised with no identifiable information.
- Representational bias – the dataset used to model social proof bias may underrepresent other consumer groups such as men since it's a women's clothing dataset.

Ethical approval was given by the University of Portsmouth (see Appendix 11.2).

1.4 Structure of Dissertation

The dissertation is structured as the following:

- Chapter 2 discusses literature on psychological biases and machine learning in consumer behaviour predictions.
- Chapter 3 outlines project planning and professional issues
- Chapter 4 outlines project requirements and research questions
- Chapter 5 reviews the datasets used with some exploratory analysis
- Chapter 6 details the methodology, including feature engineering, model training and evaluation metrics
- Chapter 7 presents the results for each bias individually as well as comparative model analysis across biases
- Chapter 8 consists of the discussion and evaluation against project requirements
- Chapter 9 provides the conclusion with a summary, reflection and future implications
- Chapter 10 consists of the references used for the literature review and datasets.
- Chapter 11 contains the appendices with an ethical approval certificate and project specification.

1.5 Chapter summary

This chapter presented the project's research challenge, aims and objectives, datasets used and justification of the study. It shed light on the gap between behavioural theory and machine learning modelling, positioning this study as an attempt to close that gap using feature engineering and interpretability (SHAP). The datasets used reflect three psychological biases grounded in behavioural economics: urgency bias, social proof and framing bias.

Chapter 2:
Literature Review

2.1 Introduction to Literature Review

This literature review will aim to explore the notion of integrating machine learning techniques with insights from decision science and behavioural economics to model and predict consumer purchase behaviour in e-commerce environments. Traditional theories in behavioural economics assume consumers are rational in their purchase behaviour, although, extensive research in psychology has implied that real-world decision making is influenced by cognitive biases, employed as marketing strategies to increase likelihood of product sales. Consumers likely deviate from rational models, particularly in e-commerce environments where there are behavioural influences, such as pricing tactics, social cues, or time constraints. Psychological theory and research suggest these deviations are predictable and driven by cognitive biases. This review focuses on three well-documented biases – urgency, social proof, and framing – all of which are currently leading systematic cognitive techniques in driving consumer purchase behaviour in modern e-commerce environments. Prospect Theory (Kahneman & Tversky, 1979) proposes individuals tend to evaluate the outcome of the decision and more sensitive to potential losses than to equivalent gains. This concept helps explain why consumers may act more impulsively to time-limited discounts, discount framing or "following the crowd" to reduce any perceived risk of purchasing less favourable products. Cialidini's (2001) Principle of Influence similarly suggests how heuristics like social proof and scarcity drive quick and subconscious purchase decisions.

This review will be organised thematically. Section 2.2 will delve into each bias, understanding psychological literature and integrating cognitive biases to structured consumer data. Section 2.3 discusses how machine learning methods have been utilised to predict consumer behaviour. Section 2.4 critically discusses the minimal research in integrating psychological theory and predictive modelling. Finally, this review will identify a research gap; although ML is brilliant in predictive modelling, underlying human decision processes are often overlooked, which this study aims address through bias-aware modelling.

2.2 Psychological Biases in Consumer-Behaviour

Where consumers act quickly towards a limited-time opportunity or fear of missing out on a product, urgency bias is at play. This behaviour roots from Prospect Theory (Kahneman & Tversky, 1979), which refers to how people are more motivated to avoid losses than to pursue equivalent gains. Such tendency is exacerbated by time pressure, causing decisions to become heuristic rather than intentional. According to Ata et al. (2021), Black Friday sales evoke higher emotional and impulsive purchase behaviour. Similarly, Raiqal & Mukaram (2025) highlighted the predictive power of urgency cues, identifying a correlation coefficient of $r = 0.744$ between seasonal discounts and impulsive purchase behaviour on a Steam platform. Godinho et al. (2016) suggests there may be a cognitive narrowing effect occurring when consumers are susceptible to time pressure, because they are more likely to buy familiar products and seek out fewer variations of that product. Additionally, Muzumdar (2021) linked impulsive purchasing intentions to a sense of urgency, particularly when it comes to promotional messaging, for instance, discount-driven holidays such as Black Friday sales, where Thomas & Peters (2011) highlight Black Friday as a "ritualised consumption event", where urgency is socially normalised. However, this material lacks ecological validity of actual transactional behaviour, given it is based on survey or experimental data.

In this project, identifying purchases made during or near significant seasonal sales periods and analysing days with abnormally high purchasing activity reflected urgency bias. Using such time-sensitive indicators help reflect when customers are most likely to experience external pressure to make rapid decisions. Although datasets frequently contain such temporal cues, machine learning models rarely interpret them as a sense of urgency – to address this, the model developed here incorporated interpretability tools (SHAP) to illustrate the impact of urgency-related factors on the model's predictions. This technique helps bridge the gap between psychological theory and model transparency.

Social proof refers to the behavioural tendency to conform to the actions of others, especially in situations of uncertainty. It is one of the six principles of influence proposed by Cialidini (2001), where popularity signals (i.e. ratings and reviews) are quick ways to gain people's trust. Social proof is most prominent in e-commerce environments, with sites such as Yelp and Amazon displaying user-generated feedback to influence purchasing decisions. ReviewTrackers (2021) helps businesses discover actionable insights about their customers, identifying that 92% of customers trust reviews as much as personal recommendations, and 94% avoid companies with negative reviews.

Muchnik et al. (2013) demonstrated a self-herding effect, where early positive votes on content skew later user responses. This shows how even without impartial assessment of the product, seeing approval of the product from others can shape collective behaviour. Through a meta-analysis, Rosario et al. (2016) demonstrated how factors that reflect popularity and perceived credibility have a significant impact on purchase decisions – such factors include the volume of reviews, how recent the review is and its sentiment. However, Askalidis & Malthouse (2016) studied selection biases in reviews, identifying consumers may not always leave feedback voluntarily, and when prompted, social expectations may affect their answers, resulting in reviews that are motivated by a wanting to conform to society norms or a positive self-image rather than genuine opinion.

This project will utilise a unique social proof score, combining average rating, review count and total helpfulness votes. Using a feature engineered variable like so reflects product reliability and influence, as well as product popularity. These variables are used by many machine learning models, but they are rarely used to represent psychological conformance. This project adds such interpretability and theoretical depth by mapping model outputs to real behavioural phenomena.

Framing bias is present when product information is presented to consumers in a manner that influences decision outcomes. Pricing strategies involve emphasising savings and discounts, even if the original value is moderate. Once again, Prospect Theory provides the theoretical foundation, where consumers show a stronger emotional response to the perceived loss of a deal than to the idea of receiving an equivalent benefit, even when the deal value remains the same. Similarly, the concept of mental accounting (Thaler, 1985) further suggests that individuals perceive gains and losses differently depending on how the information presented to them is grouped or framed. Both theories essentially suggest limited-time discounts or framed savings often compel consumers to purchase products than standard pricing, despite both offering identical prices. González et al. (2016) captured this bias, with percentage discounts yielding higher purchase intention, than fixed amount discounts due to stronger consumer perceived savings. Further empirical evidence by Zhang et al. (2013) reveals incentivised messages elicited stronger neural attention response in an event-related potential (ERP) study, particularly the P3 component, indicating there is stronger cognitive engagement in positively framed promotional messages. Other research shows that "9-ending" pricing strategies (such as £9.99) exploit left-digit bias and increased perceived value (Thomas & Morwitz, 2005, for example). However, this existing literature relies on artificial lab environments, so by employing actual e-commerce transaction data – guaranteeing ecological validity and extending the value of framing theory in computational modelling – this project enhances the applied relevance.

Featuring engineering datasets to represent framing through discount variables are used in this project. This allows the model to differentiate between price presentations that are psychologically compelling and genuine savings. While most machine learning models use discounts as numerical inputs, this project uses perception-sensitive framing cues such as discount amount, discount percentage, relative discounts compared against average category discount. Once again, the impact of these feature engineering framing traits on high-revenue purchase predictions will be quantified through SHAP analysis.

In behavioural science, all three biases – urgency, social proof, and framing – have been supported by theoretical and empirical findings. Each can be translated to psychologically significant constructs and converted into well-engineered features in structured datasets. Nevertheless, machine learning applications consider these attributes as static inputs rather than interpreting them behaviourally. By using bias-aware feature engineering along with SHAP interpretable modelling, this project fills that gap while contributing to a more human-centred approach to predictive analytics.

2.3 Machine Learning in Consumer Behaviour Prediction

A core challenge in marketing involves customer relationship management and e-commerce optimisation in predicting consumer behaviour. Given their ease of use and interpretability, traditional statistical models such as logistic regression have been used to forecast purchase intent. In a wine retail setting, for example, Van den Poel and Buckinx (2005) successfully used logistic regression to predict conversion using both demographic and transactional data – demonstrates how interpretable models can identify important determinants. Although, due to their higher predictive accuracy, more sophisticated techniques like ensemble and deep learning models have gained popularity as consumer datasets have become more complex and larger (Balyemah et al., 2024).

Ensemble methods such as Random Forest and gradient boosting (e.g. CatBoost and XGBoost) have proven to be successful. For example, in retail purchase forecasting, Lin (2025) reported *F1* scores above 0.90 and ROC AUC scores near 0.985 for XGBoost and CatBoost models, outperforming logistic regression on all criteria. Likewise, a recent comparative study captured that, when applied to consumer transaction logs, ensemble models routinely outperform single classifiers such as K-Nearest Neighbours and Support Vector Machines (Dinata et al., 2025). Thus, ensemble models can effectively handle a broad range of data, such as behavioural, temporal and environmental variables relevant to e-commerce, and they particularly excel at capturing non-linear interactions.

Deep learning models also show promise in modelling the timing and order of consumer interactions, especially sequence-aware models such as Temporal Convolutional Networks and Long Short-Term Memory (LSTM) Networks. When predicting recurrent purchased based on user-item interactions, Verma et al. (2025) discovered that time-series models performed better than tree-based models, especially in high-frequency online environments. Vallarino (2023) utilised survival analysis techniques to predict when users are most likely to make their next purchase, such as Survival Random Forests and DeepSurv. These techniques provided insights into long term customer interaction patterns, as well as predicting future purchase windows accurately.

However, a significant drawback of both ensemble and deep learning systems is their lack of interpretability. Although these models have high predictive capacity, it may be challenging to understand why a prediction was made. This presents a hurdle in consumer behaviour scenarios, especially those including behavioural theory. However, recent developments in Explainable AI have contributed to counter this. SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) are methods used to investigate the impact of various characteristics on model predictions. Research has used SHAP to reveal anchoring bias effects in Random Forest predictions, revealing that psychological price cues can be quantified and justified post hoc (Chen et al., 2023). Likewise, Azad et al (2023) trained a gradient boosting model with traits from the Theory of Planned Behaviour using LIME, they concluded how perceived control, social norms and attitude influenced purchasing decisions. These experiments emphasise that with the correct tools, interpretability can be recovered, even in complex models.

Feature engineering is also integral to success of predictive modelling. A recent case study on automated machine learning for e-commerce found that combining clickstream data from each session with long-term user attributes led to more accurate models (Dağ and Tokuç, 2025). However, few research has solely interpreted such features through a behavioural perspective. Here, current literature reveals a gap, where many models include features such as time, price and reviews, they don't tend to perceive these as psychological indicators of urgency, social influence or framing. Yixuan (2024) highlighted this problem in a review of machine learning applications in consumer behaviour, emphasising that the psychological interpretation is typically disregarded to favour the raw prediction.

2.4 Integrating Psychological Biases in Consumer Behaviour Prediction

Research is limited for integrating psychological bias theories into machine learning for consumer behaviour predictions, although recent research helps recognise its growing potential. One recent study has applied SHAP explanations to Random Forest models to demonstrate how price anchors and product features influence user clicks and engagement as a measurement of anchoring bias (Chen et al., 2023). This demonstrates that explainable AI assists in recovering behavioural impacts from machine learning models.

Bastos and Bernardes (2024) strengthen the balance between predictive power and interpretability by examining customer conversion modelling. Their study utilised gradient boosting models, incorporating SHAP and permutation importance. Findings revealed the most reliable predictors were consumer activity features and models could be interpretable without sacrificing accuracy. A recent study by Esmeli and Gokce. (2025) predicted online purchases on an e-commerce cart dataset. By using explainable AI methods, they conveyed how pricing changes and shopping/browsing session characteristics affected purchasing decisions. These examples imply explainable machine learning detects behavioural motivations, but lacks in explicitly describing psychological biases such as urgency, social proof and framing.

This notion is supported by the growth of explainable AI in general, for instance, Arrieta et al. (2019) underlined the role of XAI converting “black-box” systems – where model’s internal working and decision-making processes are not transparent such as deep learning models – into interpretable theoretical frameworks. Further research capitalises on the value of SHAP across logistic regression, tree-based models and neural networks, emphasising its ability to provide global and local explanations (Ponce-Bobadilla et al., 2024). Nevertheless, depending on SHAP can pose challenges. Feature engineering decisions particularly encoding or binning can drastically impact SHAP results, which questions its explanation stability. This is important when biases are represented through proxies or indirectly since even minor preprocessing modifications can alter which features hold most dominance and relevance.

Aside from explicitly modelling biases, Requena et al’s. (2020) research on clickstream and session data found that combining session characteristics with deep learning algorithms enhanced prediction purchase intent by still offering, and without compromising, interpretable feature importance values. Although, Tagliabue et al. (2019) stresses how complex modelling consumer sessions is, where even advanced and sophisticated systems struggle without theory-informed feature engineering. Both studies portray how machine learning in consumer contexts has predictive potential but highlights the fact that behavioural constructs are not generally treated as theoretically meaningful variables, rather as unstructured inputs.

Granted there has been such development in this field, it remains that explicit feature engineering of psychological biases to incorporate into machine learning models remains rare. Majority of consumer machine learning research consist of purchase timestamps or review counts, for example, but rarely do they attribute these variables as proxies for urgency, social proof or framing. A firm critique by Yixuan (2024) highlights this gap, arguing that machine learning tends to ignore the psychological theory behind regularly used features, restricting transparency in both theoretical and practical contexts.

2.5 Summary & Research Gap

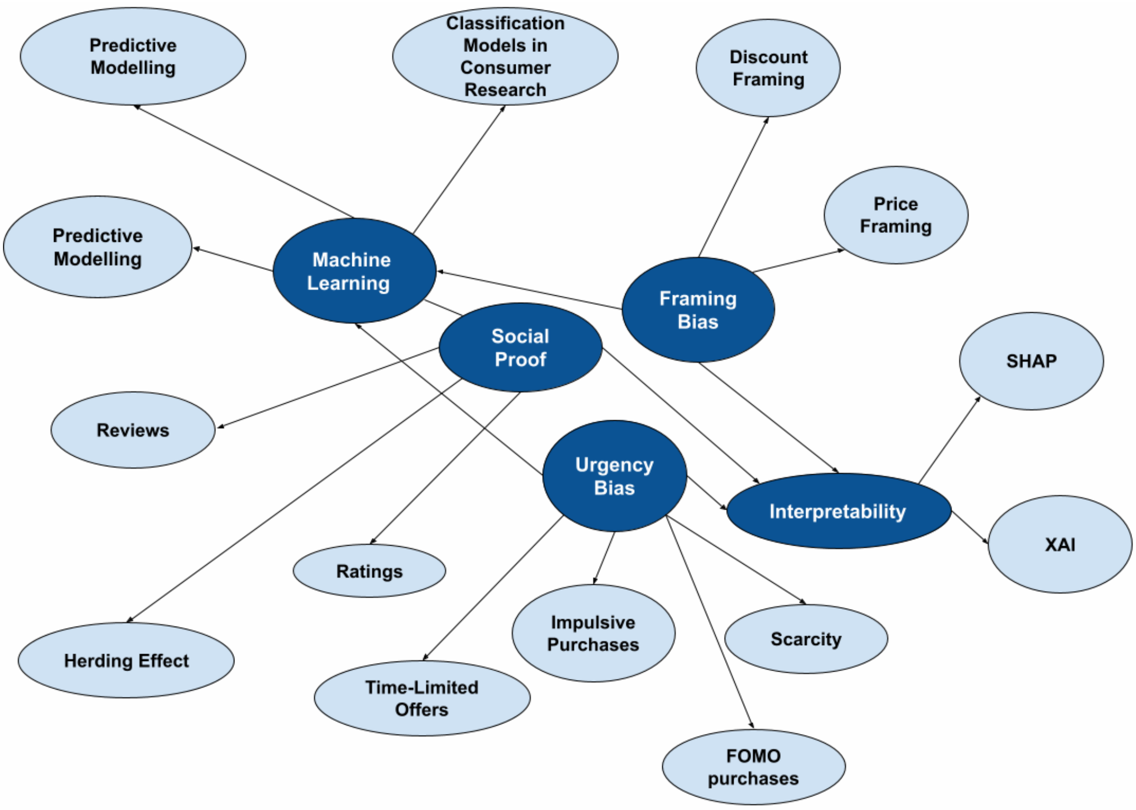
In behavioural science, all three biases – urgency, social proof, and framing – have been supported by theoretical and empirical findings. Each can be translated to psychologically significant constructs and converted into well-engineered features in structured datasets. Nevertheless, machine learning applications consider these attributes as static inputs rather than interpreting them behaviourally. By engineering features that reflect the three biases, this dissertation will address that gap and highlight how XAI can be used to represent human decision processing instead of obscuring them. Instead of depending only on typical predictors such as product category or user recency, the models used will incorporate meaningful behavioural factors that can be tracked and explained with SHAP – creating a pipeline for interpretability from theory to feature, model and explanation. Not only does this method enhance model clarity but it also aligns better with how customers make purchasing decisions, providing both academic insight and practical significance.

2.6 Search Strategy

This dissertation utilised a systematic search strategy to narrow literature from broad understanding of psychological theory and computational modelling approaches to deeper comprehension of how psychological theory can embed into machine learning models to make human decision-making processes through XAI more transparent. An initial conceptual map (Figure 1) helped structure this process, initially centring on psychological biases, interpretability and machine learning in consumer behaviour research alone, then branching out into many relevant concepts. Each section was broken down into key words and synonyms to increase breadth and coverage of literature to maximise understanding. For example, terms linked with urgency bias included scarcity, time-limited offers, fear-of-missing-out purchases, impulsive purchases; for social proof: reviews, ratings, herding effect; for framing: discount framing, price framing. Machine learning search words consisted of generic terms such as classification models, predictive modelling. SHAP and XAI was mainly used to search for literature discussing interpretability of machine learning.

Figure 1.

Conceptual Map for Literature Review Search Strategy



Databases used to locate literature included primarily Google Scholar, Science Direct, IEEE Xplore, Web of Science and the University of Portsmouth E-Library. A valuable website used was Connected Papers which allows you to explore academic literature by generating a large graph of related articles that share citations and themes from one search term. Boolean operators helped refine results (for example, “psychological biases” AND “predictive modelling”, “urgency bias” OR “scarcity”). Studies were filtered to prioritise articles between 2010-2025, but due to its theoretical groundwork, seminal work such as Kahneman and Tversky (1979) and Cialdini (2001) were included. In terms of inclusion criteria, any studies relevant to cognitive biases, consumer behaviour, machine learning, consumer behaviour and machine learning and interpretability were included. Exclusion criteria were minimal, consisting of studies only discussing psychological concepts without any AI application.

This overall strategy helped complete this literature review as efficiently as possible, providing a comprehensive and replicable approach, with a conceptual map offering transparency and structure to the search strategy.

2.7 Chapter Summary

This chapter reviewed the literature on psychological biases in consumer behaviour, machine learning, and both the concepts interacting. Three psychological biases have been recognised as top motivators in influencing consumer decision-making, but current computational models under-represent them – urgency, social proof and framing bias. Literature suggests machine learning performs well in prediction tasks, but interpretability is an issue. Using explainable AI methods such as SHAP helps address this and align machine learning algorithms in predicting consumer behaviour with psychological theory.

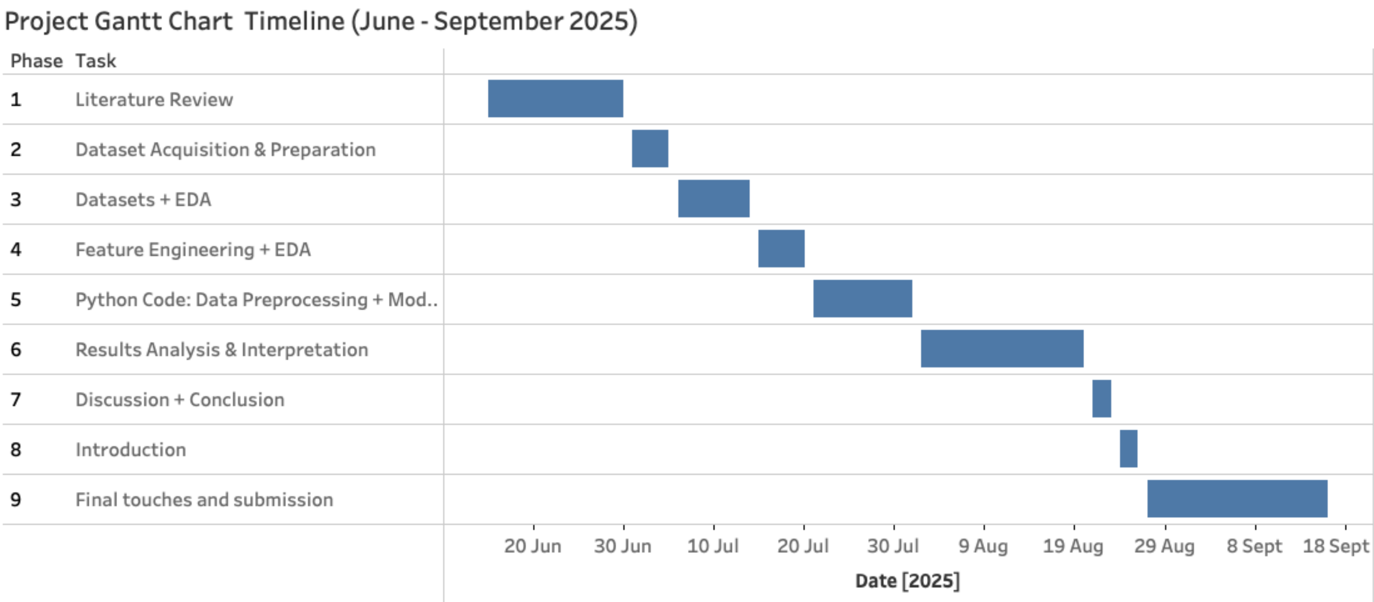
Chapter 3:

Project Planning & Professional Issues

3.1 Planning Overview

To manage this project in a time-efficient manner, effective planning was crucial. The breadth of tasks was extensive, containing a literature review, data preparation, model development, evaluation and writing up the dissertation. To make these project milestones seem more visually realistic and achievable, an initial Gantt chart was created to break this project into phases with personal set deadlines coinciding with supervisor meetings, see Figure 2.

Figure 2.
Gantt Chart of Project Milestones and Timeline



Predominantly, Google Collab was the main tool for this study, but Microsoft OneNote and Excel were used to plan dissertation structure and content and track weekly tasks to make it visually more feasible and motivating, respectively. They were very valuable tools in monitoring progress while offering flexibility to adjust personal deadlines when tasks were delayed or took longer than expected but still achieved significantly prior to the submission deadline. For instance, feature engineering proved to be me more complicated than anticipated, requiring more time and attention than planned. This led to a reallocation of time from phases which could be done quicker than the initial Gantt chart listed (i.e. exploratory data analysis). Progress was not always in line with the initial plan and sometimes required adapting the models themselves to meet deadlines, this required reflection on the efficiency of methods chosen. Models used in this project such as Random Forest and XGBoost models trained quickly within its expected timeframe, while neural networks required significantly longer, enforcing a decision to adjust parameters (reduce epochs) to speed up the process without compromising accuracy. Such scheduling signifies critical project managements and adaptability to challenges faced.

Overall, the planning process was not fixed but regularly checked and adjusted accordingly. Given the delays and adjustments required, the project demonstrates not only deadline adherence but also a deeper understanding of the challenges involved with project management on a research project of this magnitude.

3.1 Ethical, legal, professional issues

As previously mentioned, an ethical checklist in accordance with University of Portsmouth guidelines was performed and obtained official ethical approval (Appendix A). Nevertheless, given the focus on customer behaviour and psychological bias detection there were some ethical, legal and professional issues raised.

Most importantly, this project had to ethically consider data privacy and protection. All datasets utilised are publicly accessible but have no identifiable information in this report. Data governance principles such as GDPR were also considered to ensure no sensitive attributes were abused. Professional issues were raised regarding the use of open-source datasets and algorithms, creating concerns for repeatability and transparency, which was dealt with by documenting preprocessing methods in this report and citing the source of the original dataset.

By modelling cognitive biases there is a risk that findings may be misused to improve manipulative marketing, raising a further ethical issue. Although, this is mitigated since this project presents its findings around transparency and responsible AI, emphasising how bias-aware modelling can have ethical applications in supporting consumer protection rather than exploitation.

Legal considerations are flagged in using software and libraries such as Scikit-learn, TensorFlow and SHAP. Such tools are all academic-licensed and have been used within their intended scope.

All possible professional issues have been adhered too, aligning this project with the highest standard of ethical and professional practice.

3.2 Chapter Summary

This chapter discussed project planning using a Gantt chart, underlining areas of scheduling and adjustment from initial planning. It critically examined professional, legal and ethical issues ensuring the research is ethically grounded and systematically managed.

Chapter 4:
Project Requirements

4.1 Problem Specification

The key research question addressed in this project is whether psychological biases in consumer purchase behaviour can be computationally represented and recognised using machine learning. Traditional methods tend to fixate on raw predictions of purchases or revenue without investigating the behavioural underpinnings that influence these results – limiting theoretical insight and transparency

To address these challenges, five requirements were identified:

1. **Bias-aware feature engineering** – operationalising urgency, social proof, and framing as behavioural variables.
2. **Multi-model evaluation** – comparing linear, ensemble, and neural networks to assess robustness.
3. **Interpretability** – applying SHAP to test alignment with theory.
4. **Ecological validity** – acknowledging dataset limitations, particularly the use of synthetic data for framing bias.

This specification ensures the project extends beyond technical prediction to address transparency, robustness, and ethical responsibility.

4.2 Justification of Requirements

Each requirement was selected for its ability to help connect machine learning practice and psychological theory.

The key requirement for this was to use bias-aware feature engineering to ensure consumer behaviour analyses was represented meaningfully in the data. Without operationalising urgency, social proof and framing bias features, the underlying psychological influenced would remain hidden as a raw statistic. Doing such feature engineering allows models to examine if behavioural theory can in fact be embedded into predictive modelling

Multiple models were used to ensure findings were consistent and reliable across simplistic and complex algorithms. Linear models (logistic regression and SVM) are very interpretable but likely oversimplify complex behaviours. Ensemble methods (XGBoost and Random Forest) capture non-linear interactions and currently the most popular in consumer analytics. Neural networks are incredibly computationally demanding, albeit they can model more complex decision patterns. Comparison across these different algorithms provides insight into the impact of model complexity on behavioural bias prediction.

Interpretability is integral in this project. Using SHAP increases model transparency and allows models to have meaningful output by being able to trace predictions back to specific features, this helps confirm if models' decisions align with psychological theory. For instance, urgency related features should increase predicted purchases during peak sales events and holidays.

Finally, ecological validity was included as an emergent requirement. Urgency and social proof were modelled with real-world datasets, while framing relied on synthetic data. Recognising this limitation ensured the findings were contextualised, acknowledging that reliable bias detection depends not only on the algorithm but also on the quality and realism of the underlying data.

The final requirement included ecological validity since framing bias relied on synthetic data, while urgency and social proof were represented with real-world datasets. By acknowledging this limitation, it can be ensured results are further contextualised, recognising that data must be of good quality and realism for reliable bias detection.

4.3 Research Questions

From these requirements, three research questions were formulated:

1. *Can machine learning models detect psychological biases when engineered into datasets?*
   - Tests the feasibility of computationally representing psychological constructs.

2. *Which models (linear, ensemble, neural) perform most effectively across biases?*
   - Evaluates whether model complexity contributes meaningfully to behavioural prediction.

3. *What features most strongly drive model predictions, and do they align with psychological expectations?*
   - Links interpretability to theory testing, identifying convergence or conflict between computational and behavioural perspectives.

The project requirements were systematically mapped against the research questions to ensure alignment. Table 1 links each requirement to its associated research question.

Table 1.
Requirements to Research Questions Traceability Matrix

Requirement	Linked Research Question(s)	Rationale
Bias-aware feature engineering	RQ1, RQ3	Examines if all three biases can be represented through engineered features from raw data and if machine learning models can present them in a way aligning with psychological theory.
Multi-model evaluation	RQ2	To make sure results are not specific to models by comparing across linear, ensemble and neural network algorithms. Further demonstrates how model complexity has effect in capturing behavioural influences.
Interpretability (SHAP)	RQ3	Traces predictions back to features, providing transparency, confirming if feature importance represents psychological expectations
Ecological validity	RQ1	To ensure results are contextualised by recognising dataset limitations (using real-world datasets versus a synthetic one), reflecting on the impact of data quality on reliability.

4.4 Chapter Summary

This chapter discussed project requirements and their importance, as well as the project's research questions. It provided a cohesive framework for detecting consumer decisions by linking feature engineering, multi-model evaluation, interpretability, and ecological validity with the study objectives and requirements.

Chapter 5:

Data Collection & Exploratory Data Analytics

5.1 Urgency Bias Dataset

To model the tendency to make quicker decisions under time pressure, this project utilised a real-world transactional dataset consisting of Amazon purchases, sourced from Harvard Dataverse. It consists of 1,850,720 rows of customer purchase records collected between 2018-2022 in US Regions. Each row refers to individual purchase records, based on fields such as order timestamps, unit price, quantity etc, see Table 2.

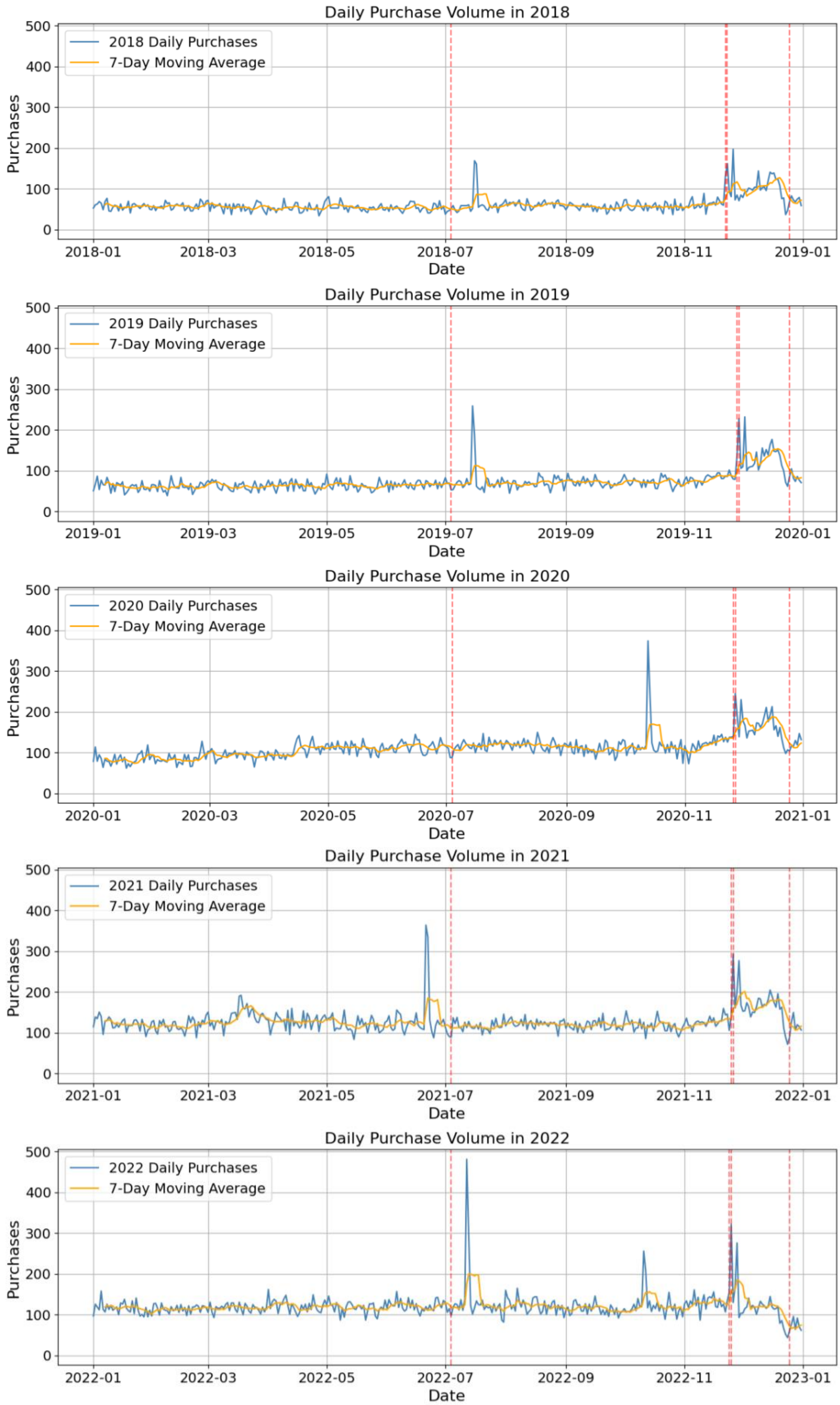
Table 2.
Sample of raw urgency bias dataset

Order Date	Purchase Price Per Unit	Quantity	Shipping Address State	Title	ASIN/IS (Product Code)	Category	Survey ResponseID
2020-01-12	13.98	1	NY	Huggies Natural Care Sensitive Baby Wipes, Unscented, 48 Count (Pack of 6)	B07M6 XVCBY	SKIN_ CLEANING_ WIPE	R_2uyQHw ENySBamVs
2018-01-27	7.98	1	PA		B008 ZZC6S6		R_3k5Ja bbpW5kTO7M
2020-05-20	6.27	1	PA	Amazon Basics RJ45 Cat-6 Ethernet Patch Internet Cable - 5 Foot (1.5 Meters)	B00N2VILDM	COMPUTER _ADD_ON	R_2QYCIl75 oJ8y4OX
2021-11-29	16.24	1	TX	Dreams - PlayStation 4	B082LW28B2	PHYSICAL_ VIDEO_ GAME_ SOFTWARE	R_31pjiIS 2z0L0g9l
2021-01-17	29.39	1	NV	Craig Frames 200ASH Poster Frame, 20 x 27, Black	B0049OCFZ6	PICTURE_ FRAME	R_3GrjtiJ9 M19hViM

This dataset is relevant to consumer decision-making patterns, deeming it suitable for this study, enabling the inference of behavioural indicators tied to psychological urgency. The raw data alone may not signify urgency or psychological annotations, although its structure and foundation support later feature engineering to investigate time-based trends, i.e., whether there are spikes in orders around holidays. For instance, in this dataset, daily purchase volume from 2018-2022 spiked during seasonal events, see Figure 3. Clear volume surges arise at the end of each year between November and December, aligning with seasonal sales events such as Black Friday, Christmas and Thanksgiving while mid-year peaks in July reflect flash sales or summer sales. A 7-day moving average was used to emphasise these seasonal patterns, demonstrating that consumer purchases are urgency driven by events. This supports the purpose of later urgency-related feature engineering such as "Days to Nearest Holiday" and "Holiday Proximity."

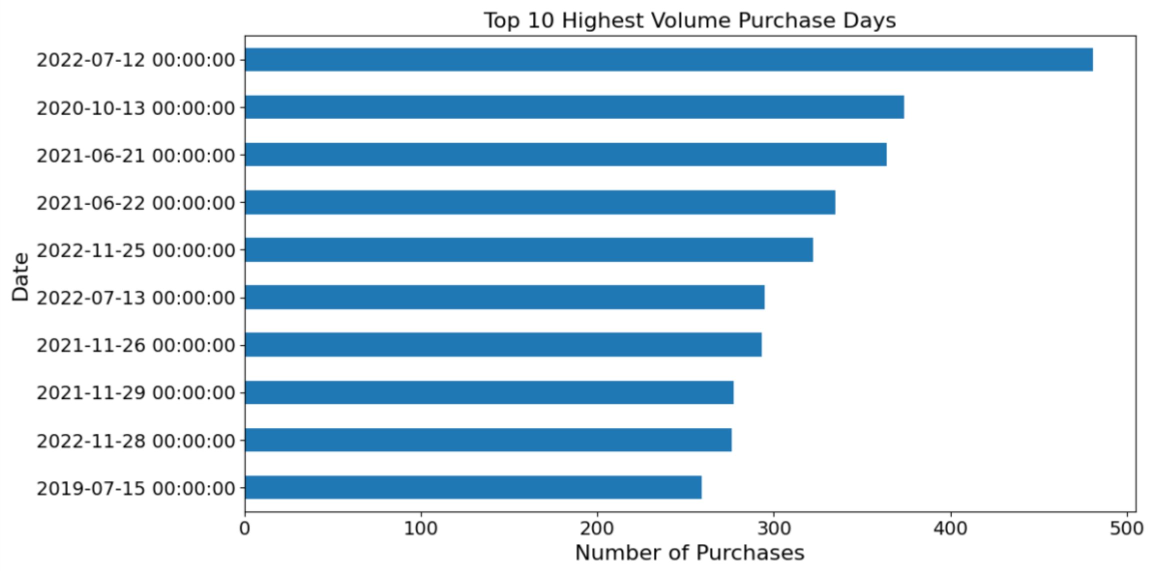
Figure 3.

Time-Series Figure showing daily purchase volumes from 2018 to 2022 with a 7-day moving average



Similarly, the top 10 highest-volume purchase days between 2019 and 2022 also cluster around mid-year flash events and sales as well as November periods, once again highlighting the urgency-driven behaviour in consumer purchase behaviour, see Figure 4.

Figure 4.
Horizontal Bar Chart of the Top 10 Highest Volume Purchase Days



5.2 Social Proof Bias Dataset

Social Proof was modelled using a dataset about Women's Clothing E-Commerce Reviews, consisting of 23,486 user-generated reviews on clothing products. Again, this dataset was sourced from an accessible repository on Kaggle, it captures a rich account of both qualitative and quantitative cues that impact consumer purchase decisions in online retail environments. Each row refers to individual product purchases per user, see Table 3.

Table 3.
Sample of raw social proof dataset

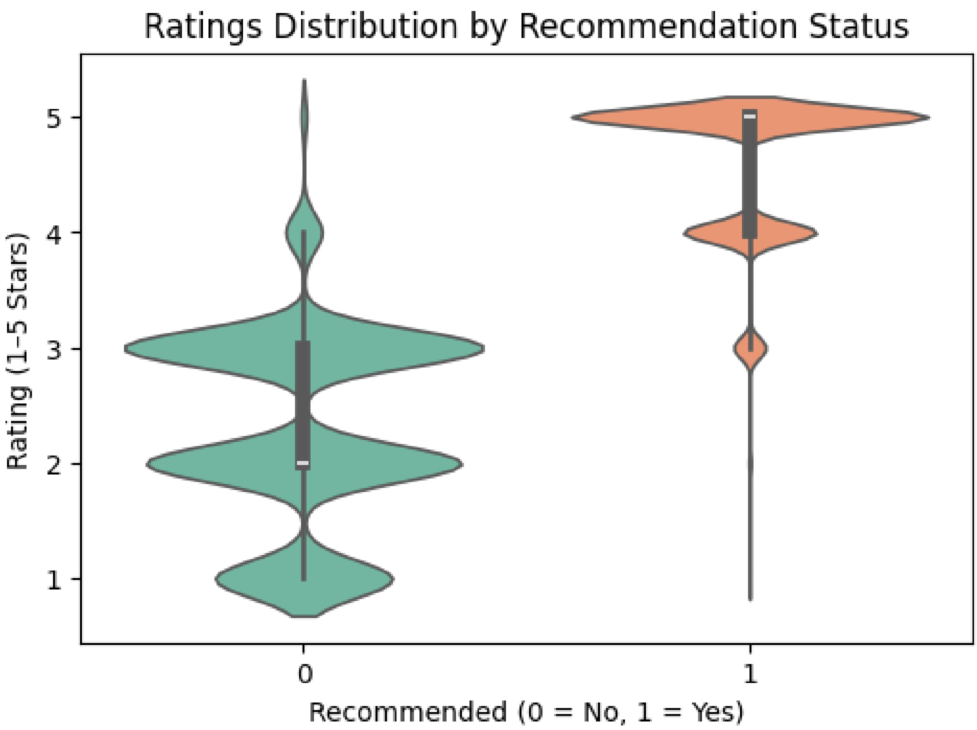
Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name
767	33		Absolutely wonderful - silky and sexy and comfortable	4	1	0	Intimates	Intimate	Intimates
1095	39		This dress is perfection! so pretty and flattering.	5	1	2	General Petite	Dresses	Dresses
1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, flirty, and fabulous! every time i wear it, i get nothing but great compliments!	5	1	0	General Petite	Bottoms	Pants

Combined, these variables can be considered proxies for social proof cues, providing insight into product popularity through peer endorsement (recommendations), review quality (feedback count) and sentiment of review (rating). Qualitative aspects of open-text reviews allow further analysis of language, tone and sentiment. Given these variables, this dataset is suitable to assess these biases influence on purchase behaviour as we can navigate and model how individuals align their purchase behaviour with social norms or peer approval.

Ratings were analysed in connection to the recommendation outcome, suggesting that products are grouped around higher star ratings while products not recommended exhibit lower representation values (Figure 5). The violin plot demonstrates that reviews labelled as not recommended (0) are more widespread across lower rating values (1-3), while recommended reviews are visibly skewed towards higher rating values (4 and 5) with little to no representation of lower ratings. This strengthens the notion that higher ratings act as social proof signals that strongly influence purchase decisions.

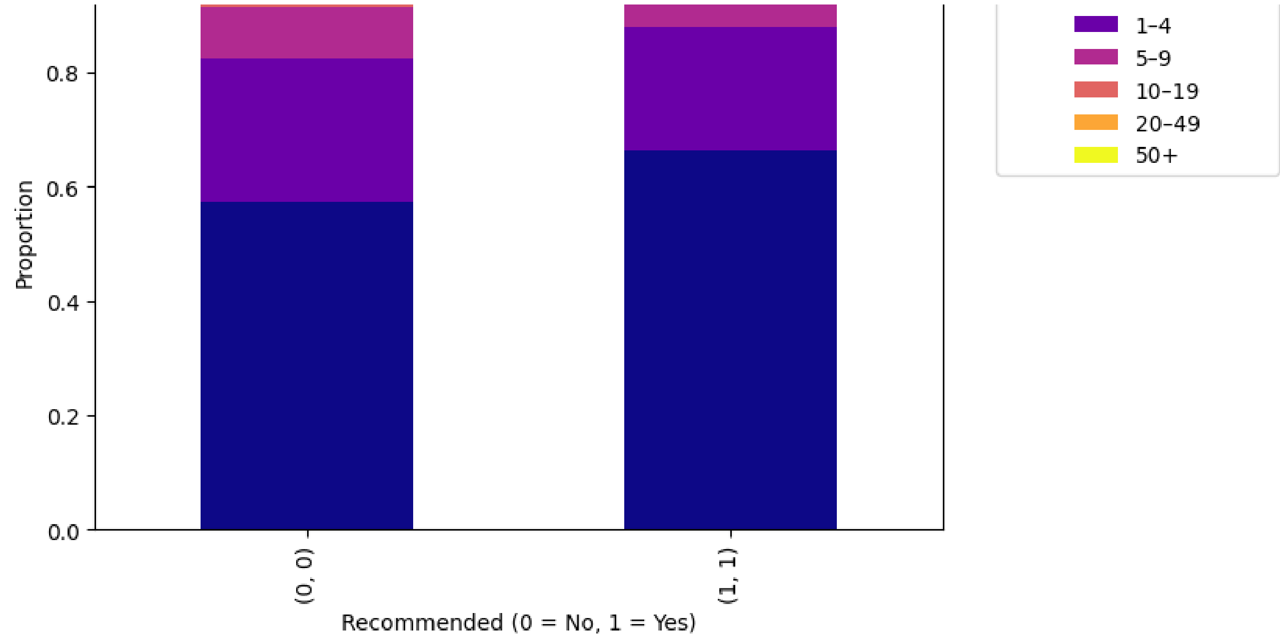
Figure 5.

Violin Plot showing the distribution of Ratings (1-5) by Recommendation Status (0=Not Recommended, 1=Recommended)



Positive feedback totals the number of endorsements from other users to individual reviews, providing another feature of social validation. By examining its distribution against recommendation class, it can be inferred if recommended items attract more attention from other users, reinforcing the role of social proof cues in influencing consumer purchase behaviour. The proportion plot below (Figure 6) reveals majority of reviews received no positive feedback, regardless of the recommendation class. Although, where reviews had more feedback, recommended items have a greater proportion of the 5-9, 10-19, and 20+ positive feedback counts. This implies positive feedback attracts more peer validation providing a complementary but noisy feature of social proof influence.

Figure 6.
Proportion Chart showing Positive Feedback Count (Bins=6) by Recommendation Status (0=Not Recommended, 1=Recommended)



5.3 Framing Bias Dataset

To investigate how individuals' cognition is distorted when having to make different choices based on how the information is presented (framing bias), a synthetic e-commerce dataset sourced from Kaggle was used. It contains 100,000 purchase records across multiple product categories and regions. Key variables include discount applied, units sold, revenue made and advertisement engagement data, see Table 4.

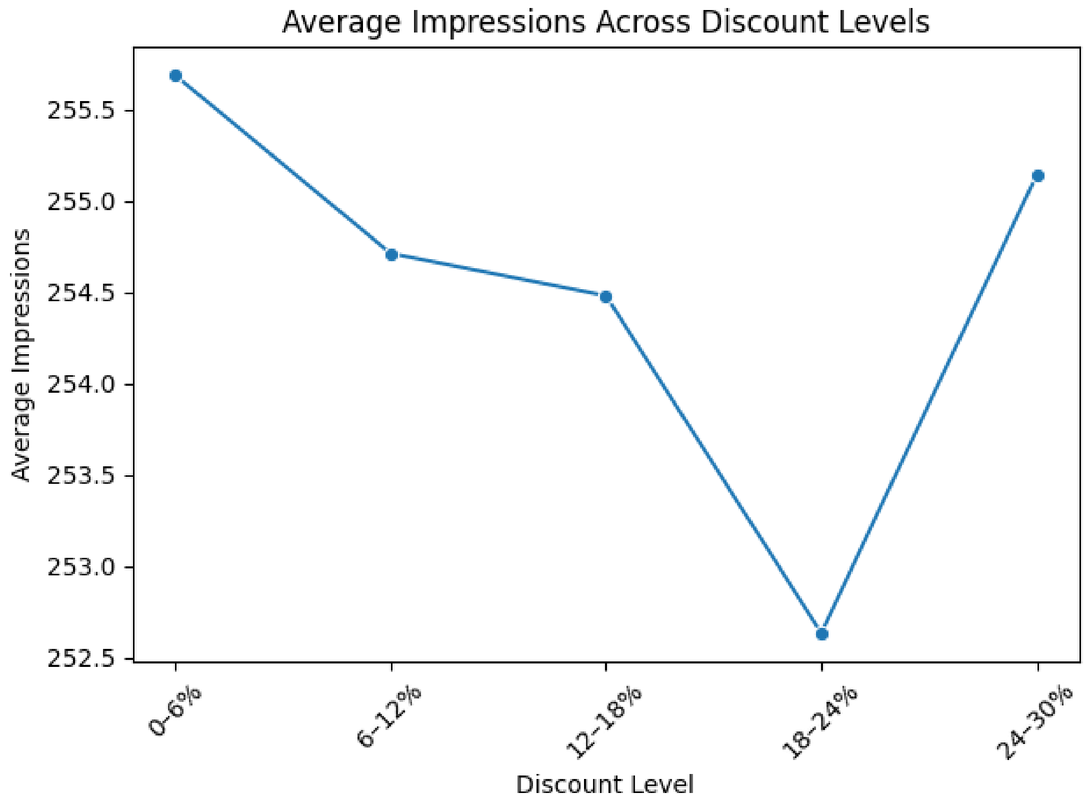
Table 4.
Sample of raw framing bias dataset

Customer_ID	**Product_ID**	**Transaction_Date**	**Units_ Sold**	**Discount_ Applied**	**Revenue**	**Clicks**	**Impressions**
Customer_65	Product_224	2024-10-06	134	0.14	305.54	11	65
Customer_1910	Product_584	2024-10-29	109	0.3	1102.19	15	201
Customer_2306	Product_374	2024-04-04	116	0.04	471.29	16	199
Customer_17206	Product_220	2024-08-25	125	0.2	980.26	12	355
Customer_16033	Product_358	2024-05-05	132	0.07	803.76	44	355
Conversion_Rate	**Category**	**Region**	**Ad_CTR**	**Ad_CPC**	**Ad_Spend**		
0.17	Electronics	Europe	0.018	0.55	9.9		
0.07	Home Appliances	Asia	0.1589	0.4	63.56		
0.08	Toys	Asia	0.0596	1.5	89.4		
0.03	Clothing	Europe	0.0444	0.44	19.54		
0.12	Books	North America	0.127	0.53	67.31		

These characteristics both directly and indirectly frame how value and urgency can be shaped, for example, by repeatedly exposing individuals to advertisements, it may reframe the appeal of a product, as well as exposure to an initially high price followed by a decrease may reframe product desirability. Therefore, this dataset is very suitable for feature engineering using psychological framing constructs.

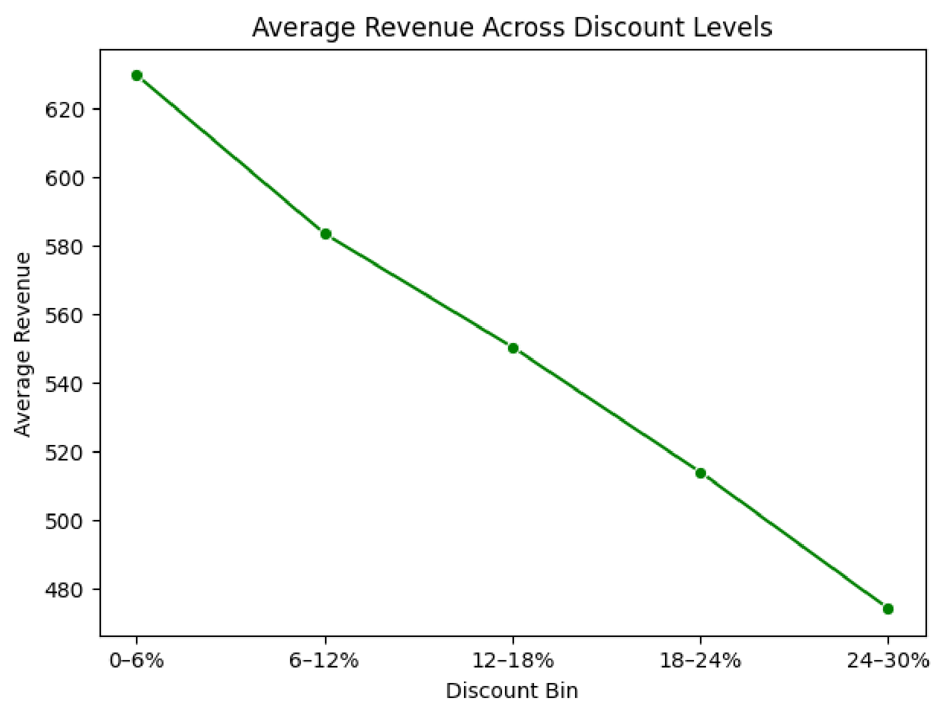
To investigate the relationship between discounting and consumer exposure, impressions were compared against discount levels. Unexpectedly, products with low or minimal discount obtained the most impressions while those with moderate discount (18-24%) received lower impression levels (Figure 7). This implies visibility was effects by other factors external to the raw dataset, such as category popularity or advertising. Discounts seem to have less impact in boosting exposure and thus influencing purchase behaviour, which reinforces the need for feature engineering in untangling the effect of framing bias.

Figure 7.
Line Chart of Average Impressions of Products against Discount Levels



To see how discounts affect sales revenue, average revenue was compared across discount levels to identify if larger discounts impact revenue patterns. Figure 8 shows a downward trend of decreasing average revenue as discounts increase. Where products had little to no discount, it obtained the highest revenue while the highest discounted items had the lowest revenue. This suggests discounts may cause product value to be perceived differently but not always lead to sales success.

Figure 8.
Line Chart of Average Revenue against Discount Levels



All datasets were anonymised and do not include any personally identifiable information. They are all publicly accessible and used for academic research purposes only. It has been ethically approved by the University of Portsmouth ethics board (see Appendix 11.2).

5.4 Chapter Summary

This chapter outlined and explored the three datasets used to separately model urgency, social proof and framing biases. Exploratory data analysis uncovered temporal spikes in consumer purchases, revealing urgency bias, review patterns and recommendations reflected social proof and discount effects mirrored framing bias.

Chapter 6:

Methodology

6.1 Data Preprocessing

Different preprocessing steps were conducted depending on the size and quality of the datasets. For both urgency bias and social proof datasets, rows with null values were eliminated to address missing data. This is necessary as missing data prohibits classification models from performing. Given that the proportion of missing values was small relative to the total dataset size and that there was no obvious pattern of bias in the deleted values, this was a pragmatic decision. In contrast, the dataset used to model framing bias was complete, requiring no row removal, all 100,000 transactions were kept after a validation check for completeness.

Additionally, the urgency bias dataset exceeds one million records, so a 10% stratified sample was taken to drastically reduce training time while maintaining class distribution and statistical representativeness, for the change in shape of each dataset see Table 5.

Table 5.
Dataset Transformation

Dataset	Original Rows	Final Rows	Columns
Urgency Bias	1,850,720	167,534	8
Social Proof Bias	23,486	19,662	11
Framing Bias	100,000	100,000	15

At this stage, there were no further transformations or cleaning steps, each dataset was retained in this state prior to beginning of feature engineering and no columns were removed. After preprocessing, each dataset was split into training and testing subsets (80/20% stratified split) to ensure class balance across target variables. This was crucial to make sure model evaluation is fair and consistent during cross-comparison between different models and biases.

6.2 Feature Engineering

To reflect three psychological biases, it was essential for this study to engineer features from raw behavioural data seen in the previous section. Since these biases were not directly observable in the raw data, features were engineered to behave as proxies for these biases. This was completed separately for each dataset and its targeted bias.

6.2.1 Urgency Bias Feature Engineering + EDA

As Cialdini (2007) mentioned, urgency bias is present where time pressure or scarcity instigates quicker consumer purchase decisions. Three key features were used to represent whether a purchase was urgent or not characterised by the *'urgency label'*, see Table 6 for a summary of the feature engineering process.

Table 6.
Summary of Feature Engineering the Urgency Bias Dataset

Feature	Creation	Justification
Holiday Purchase	The first binary urgency trigger which cross-referenced each purchase date of transactions with manually defined USA holiday dates from 2018-2022 (Black Friday, Thanksgiving, Independence Day and Christmas)	Consumers compete for inventory during time-limited deals (Chevalier & Mayzlin, 2006). They fear the necessary inventory and essentials will be out of stock, triggering a sense of purchase urgency around holidays.
Days to Nearest Holiday	Counts the number of days between the purchase data of each transaction and the closest manually defined holiday.	Time-sensitive purchases aren't made on the day of holidays but most likely within a few days leading up to it (Grewal et al., 1998), this captures other holiday-related high-volume purchases.
Purchase Amount and High Value Purchase	"Purchase Amount" is calculated by multiplying unit price and quantity. Of these transactions, the "High Value Purchase Variable" formulates the top 10% and is the second binary urgency trigger.	Loss aversion is heightened with limited time discounts and accelerates high-value purchases (Aggarwall & Vaidyanthan, 2003)
Daily Volume and High-Volume Day	"Daily Volume" calculated total purchases on a given date, and the "High Volume Day" binary variable flags date in the top 10% of this distribution and the third binary urgency trigger.	In instances of scarcity and popularity cues such as low stock or trending status, consumers tend to perceive this as an urgent purchase and more likely to make bulk or impulsive purchases (Li et al., 2020).
Urgency Label	Binary target label (1 = urgent purchase, 0 = not urgent) Created using OR logic across (holiday purchase, high value purchase, and high-volume day variables)	Assumes a single urgency trigger is enough to suggest transactions were influenced by sense of urgency. This may seem like a sensitive approach, but research suggests that even one time-pressure cue can impact consumer purchase behaviour (Monroe & Suri, 2003).

To ensure machine learning readiness, all qualitative columns were removed while retaining numerical and feature engineered variables, see Table 7.

Table 7.
Machine Learning Ready Dataset for Predicting Urgency Bias-Induced Purchases

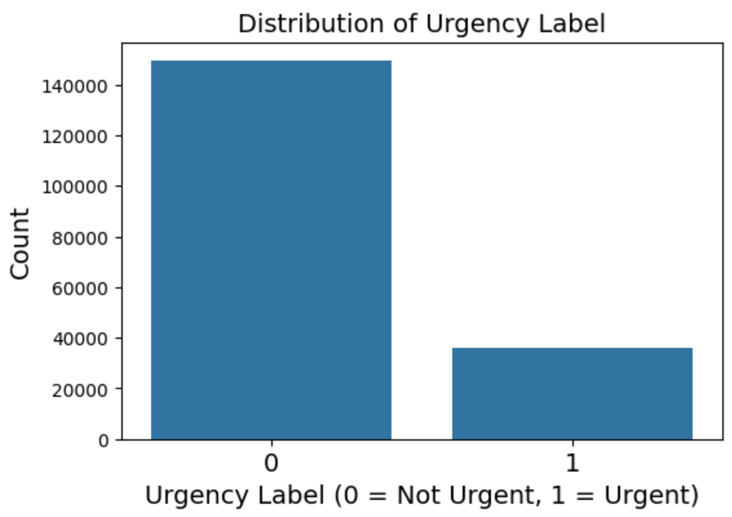
Order Date	Purchase Amount	Holiday Purchase	Days to Nearest Holiday	High Value Purchase	High Volume Day	Urgency Label
2021-02-06	9.99	0	43	0	0	0
2020-12-23	108.99	1	2	1	1	1
2018-01-27	7.98	0	299	0	0	0
2018-10-09	15.99	0	44	0	0	0
2019-05-14	22.99	0	140	0	0	0

Exploratory Data Analysis

The distribution of the target variable **(**Figure 9) indicates that approximately 20% of transactions are urgency induced, which is psychologically plausible given urgency is not an everyday occurrence. Using OR logic across the three binary triggers yielded a reasonable proportion of urgency cases. Classes may seem imbalanced but balancing it in this case takes away from behavioural realism and risks overfitting.

Figure 9.

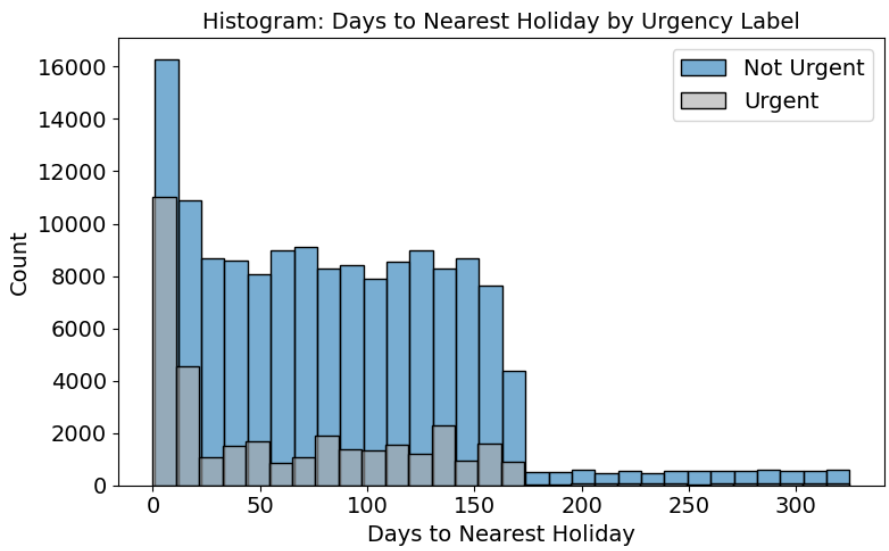
Bar Chart: Distribution of Urgency Label



Further feature distribution analysis supported the predictive value of the feature engineered attributes. The histogram **(**Figure 10) demonstrates that purchases identified as urgent cluster within 0-20 days of national holidays, which satisfies the psychological expectation of time-sensitive deals inducing decision-making.

Figure 10.

Histogram: Days to Nearest Holiday by Urgency Label



Limitations:

Although these features help reflect urgency behaviour, there are drawbacks. It is likely some purchases made on holidays are not urgency-driven but the 'holiday purchase' feature includes these as a representation of urgency. Additionally, the 'daily volume' may capture high volume days that are due to other promotional events not accounted for in this dataset. Regardless, these features in general provide a fully operational framework for identifying urgency-related purchase behaviour in e-commerce transactions.

6.2.2 Social Proof Feature Engineering + EDA

To transform raw review data into indicators of social proof, features were engineered to establish behavioural cues that psychological theory has proven to influence consumer decisions, including perceived quality, social endorsement and popularity of items, see Table 8.

Table 8.

Summary of Feature Engineering the Social Proof Dataset

Feature	Creation	Justification
Average Rating Per Item	The mean Rating for each Clothing ID	Higher ratings persuade potential buyers into believing the quality of the product is of a high standard which influences purchase decisions even if they do not even need that product (Cialdini, 2007). This feature was created to reflect the quality of each product based on aggregated customer ratings.
Review Count Per Item	Total number of reviews for each Clothing ID	High review counts indicate products are in trend and popular, a greater number of reviews often gain consumer's trust and appeal, increasing the likelihood of a purchase (ReviewTrackers, 2023)
Total Helpfulness Per Item	Total "Positive Feedback Count" per Clothing ID compute the combined helpfulness of reviews for each item	Social validation is another factor that gains trust and appeal from customers. (Chevalier & Mayzlin, 2006). So, analysing social validation in terms of helpfulness scores indicates the strength of community agreement on purchase influence.
Social Proof Score	Totalling the three indicators above.	A singular binary variable reflecting overall social endorsement of items. Combines social cues that imitate real-world pathways to decision-making: quality (rating), quantity (review count), and influence (helpfulness).
Recommended IND (target)	Not a feature engineered variable as it was taken directly from the original dataset, indicating if a product was either recommended (1) or not (0).	Acts as the target variable to predict its outcome based on social proof features.

To produce a machine learning ready dataset, the final structure reduced to a minimal but information-rich version in comparison to the original dataset, see Table 9.

Table 9.

Machine Learning Ready Dataset for Predicting Social Proof Bias-Induced Purchases

Clothing ID	Average Rating Per Item	Review Count Per Item	Total Helpfulness Per Item	Social Proof Score	Recommended IND
767	4.5	2	0	0.875978	1
1080	4.294118	289	849	1.412329	1
1077	4.084175	297	702	1.31446	0
1049	4.3125	32	90	0.891001	1
847	4	4	8	0.755828	1

Exploratory Data Analysis

Figure 11.

Histogram representing distribution of Social Proof Scores

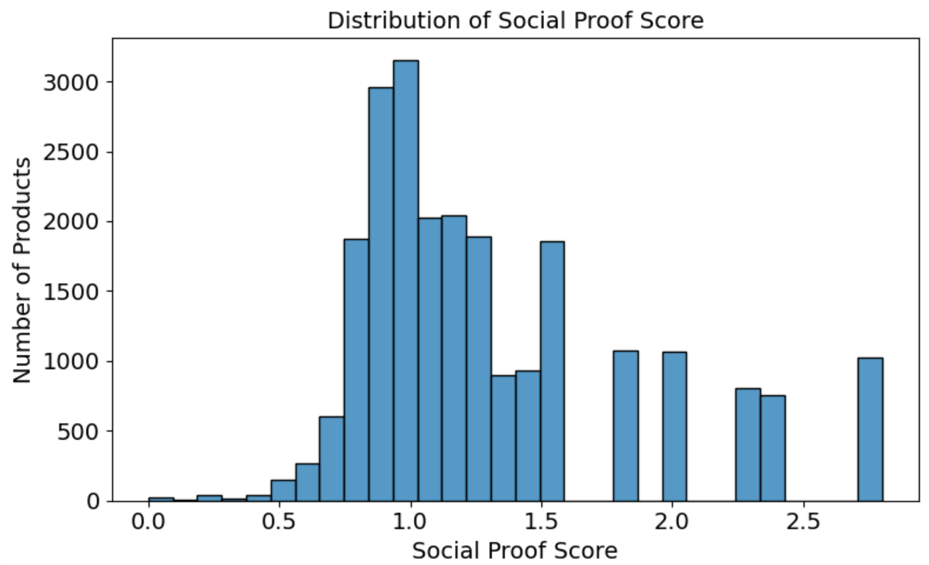


Figure 11 captures the distribution is slightly right-skewed, with majority of scores ranging between 0.8 and 1.3 – suggesting most items have moderate social approval. Fewer products showcased higher social proof scores, indicating strong endorsement. Low-scoring outliers indicate that less-reviewed or more recent items were included in this dataset. Overall, the distribution in social proof scores can capture the ability of feature's to meaningfully distinguish between different degrees of social influence in consumer behaviour.

Figure 12.

Boxplot representing the distribution of social proof scores by recommendation status

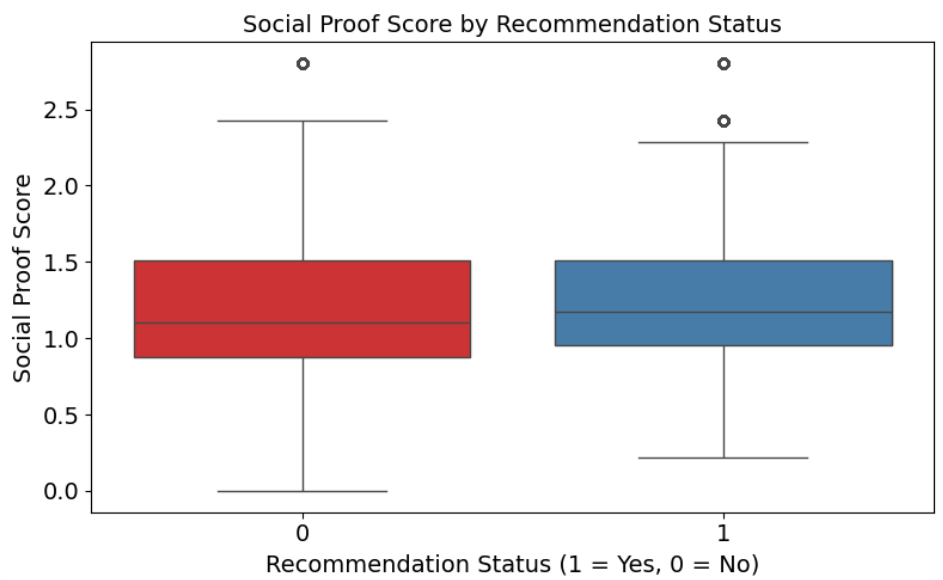


Figure 12 demonstrates how products that were recommended are only marginally higher in social proof scores on average (median), and their scores are more closely clustered (interquartile range). So, products are more likely to be recommended with stronger social signals such as more helpful feedback, review or higher ratings. There is substantial overlap between both status variables, implying other factors may influence a person's decision to recommend a product.

Limitations

Firstly, aggregating review data operates under the assumption that features such as ratings, reviews and helpfulness scores and presented to customers equally, which may not be true in actual online browsing experiences. Additionally, the target variable "Recommended IND" does not account for personal factors and could be influenced by aspects such as fit or style preferences – adding noise to the research. Thirdly, the social proof scores give equal weight to helpfulness, review count and rating, but customers may prioritise these social proof signals differently than this method proposes. Also, ratings are reviews are fundamentally subjective, producing inconsistencies across users and products; similarly, popularity bias further skews outcomes with products of high visibility receiving more reviews than niche items, which furthers adds noise in the dataset and limits the explanatory power of social proof predictors. A more sophisticated model could assign different weights to these cues, relative to their importance.

6.2.3 Framing Feature Engineering + EDA

To model how discount framing influences purchase decisions, features were engineered from the raw dataset to isolate discount variables to better capture its influence in purchase likelihood.

Table 10.
Summary of Feature Engineered Process for Framing Bias

Feature	Creation	Justification
Average Unit Price	Calculated using variables from the initial dataset: 'Revenue' divided by 'Units Sold'	Provides the price per unit so discount effects across purchases of different sizes can be compared.
Discount Amount	Average Unit Price multiplied by "Discount Applied" (from the initial dataset)	Calculates discount amount per unit – capturing the absolute savings a consumer sees.
Category-Level Average Discount	The mean discount included in each product category	Represents the perceived value influenced by category norms, providing the contextual foundation for framing effects – i.e. 'is this a better deal than usual.'
Relative Discount	Category Average Discount multiplied by Discount Applied (initial dataset)	Quantifies the framing bias. People may feel more compelled to purchase products if an items discount is bigger than average.
High Revenue Target	Binary variable where revenue is either of high revenue or greater than the calculated median value, (1), otherwise it is considered low revenue (0).	Uses 'Revenue' as a proxy for strong purchase uptake, and acts as the label for supervised learning.

Many columns in this instance were irrelevant and dropped, leaving the final dataset ready for machine learning, see Table 11.

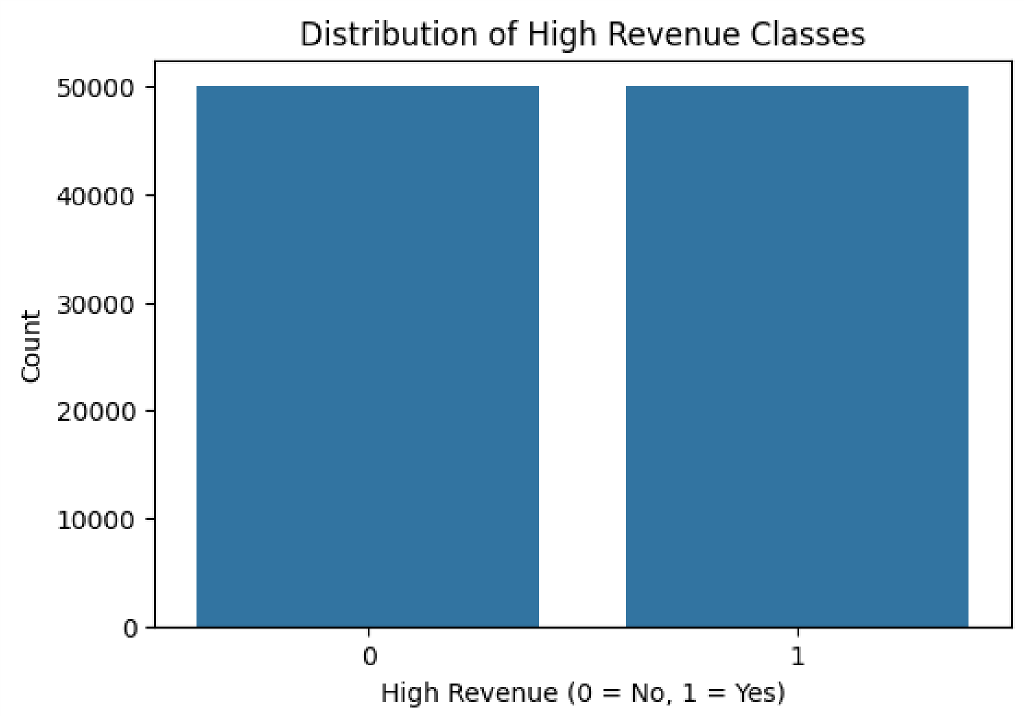
Table 11.
Machine Learning Ready Dataset for Framing Bias-Induced Purchases

Category	Units Sold	Discount Applied	Average Price	Discount Amount	Category Average Discount	Relative Discount	High Revenue
Electronics	134	0.14	2.280149	0.319221	0.150005	-0.01001	0
Home Appliances	109	0.3	10.11183	3.03355	0.149167	0.150833	1
Toys	116	0.04	4.062845	0.162514	0.149679	-0.10968	1
Clothing	125	0.2	7.84208	1.568416	0.149745	0.050255	1
Books	132	0.07	6.089091	0.426236	0.150779	-0.08078	1

Exploratory Data Analysis

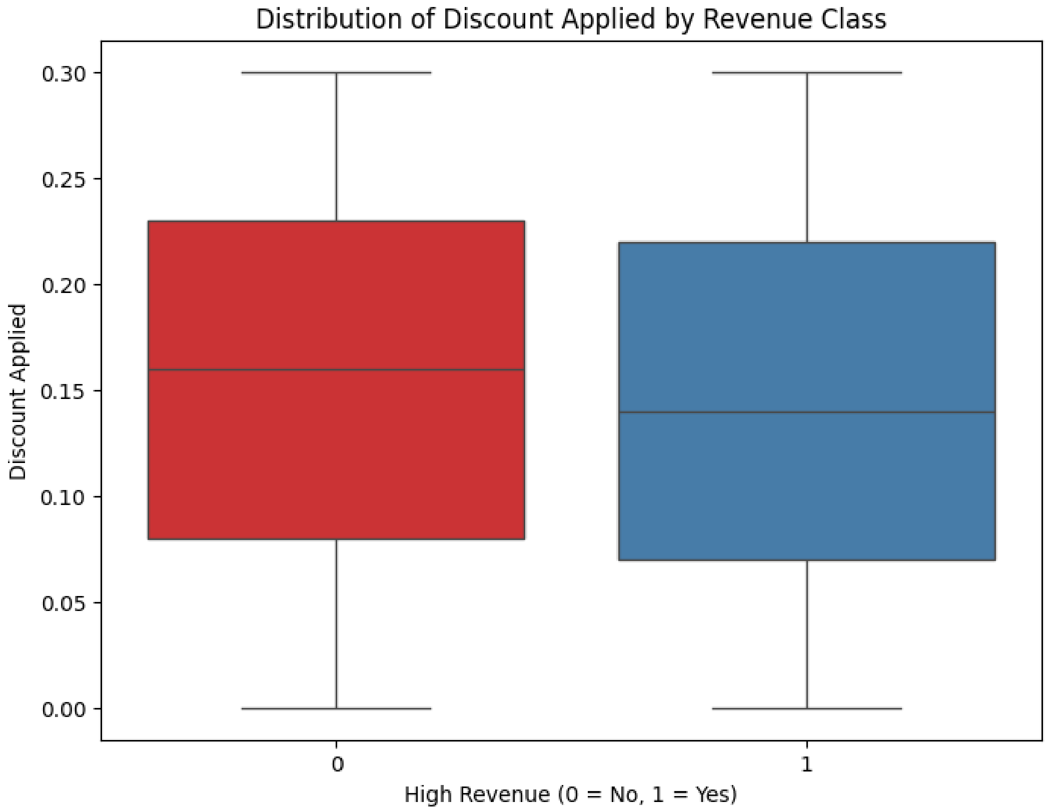
Firstly, the distribution of the target variable was examined. Figure 13 shows the dataset is perfectly balanced with an equal split between low revenue (0) and high revenue (1) cases. This even distribution stems from the engineering of the High Revenue variable where transactions were grouped as a high revenue purchase if they were above the median value and not high revenue if they were below it. This balance reduces the risk of bias and skewness of results.

Figure 13.
Bar chart showing the distribution of the High Revenue Classes



To investigate if there is an effect of discounting on purchase behaviour, a boxplot was created to compare discount values across revenue classes. Figure 14 demonstrates that both groups are influenced by discounted prices with the 'not high revenue' class having a marginally higher median and larger variability. An overlap in distributions indicates discount value is not the main cause of high-revenue classification.

Figure 14.
Boxplot comparing discount levels across revenue outcome



These findings provide a neutral framework for later modelling of how discount framing effects impact consumer decisions.

Limitations

This feature engineering approach does have some drawbacks. Discount-related features, including Discount Applied, Relative Discount and Category Average are very similar which raises a concern of multicollinearity which might make the contribution of individual predictors obscure. Secondly, creating a target variable based on the median imposes an artificial barrier which oversimplifies what is fundamentally a continuous variable. Thirdly, discounts may capture some aspects of framing bias but they were not able to fully represent the psychological factors that influence consumer decisions, hence the use of proxies.

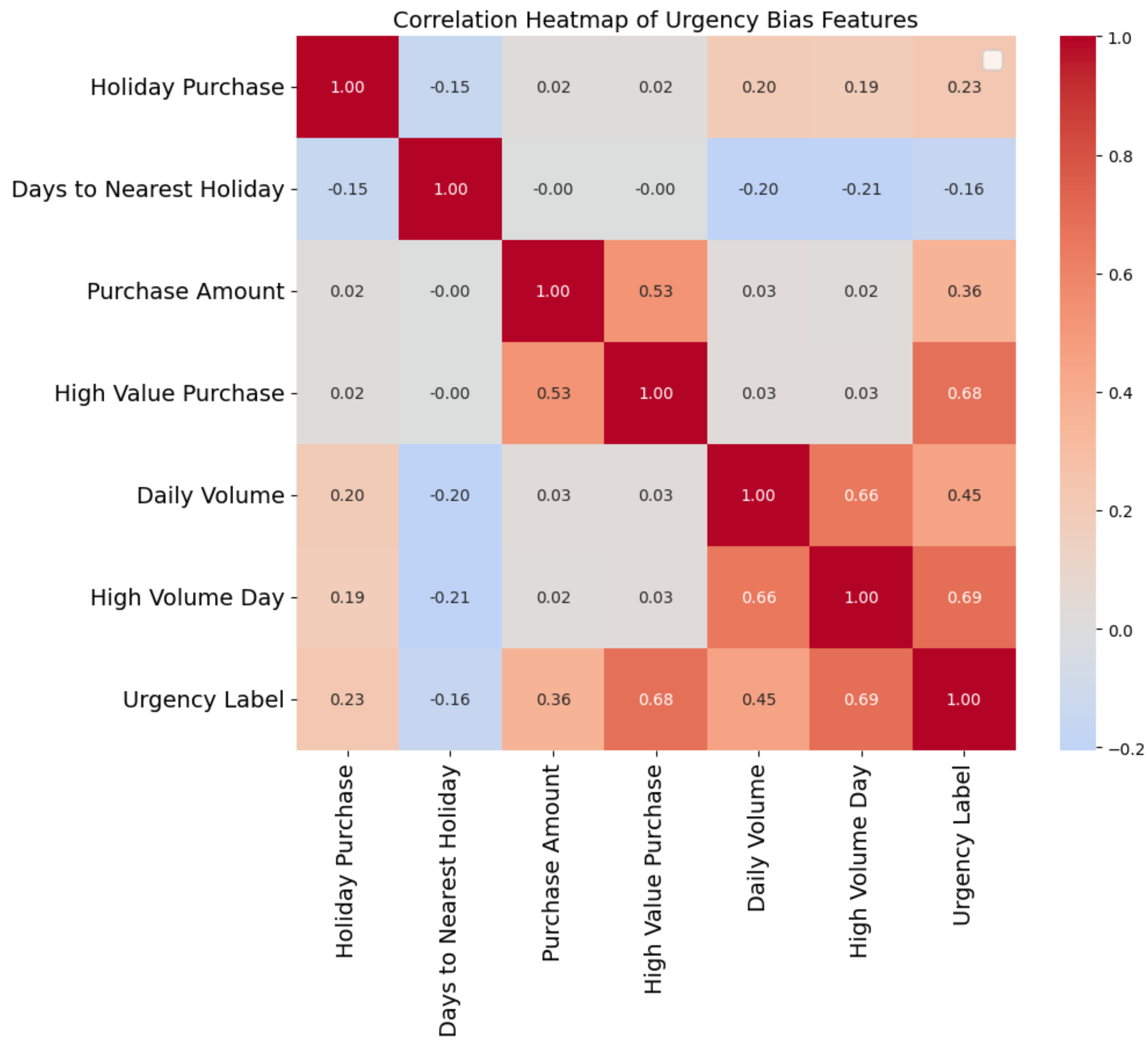
6.3 Exploratory Correlation Analysis

Correlation analysis was conducted to investigate the relationships between the engineered features, particularly against the target variable across all three bias-specific datasets. To visualise the relationships, heatmaps were produced (Figures 15 - 17).

In the feature engineered urgency bias dataset, the target 'Urgency Label' had stronger correlations with variables related to transaction size and time. For instance, 'High Value Purchase' and 'High Volume Day' show stronger positive correlation with the target variable, while features such as "Days to Nearest Holiday" have a weaker correlation. This implies that urgent instances are more closely associated to high-value, high-volume activity rather than holiday calendar proximity.

Figure 15.

Correlation Heatmap of Urgency Bias Features

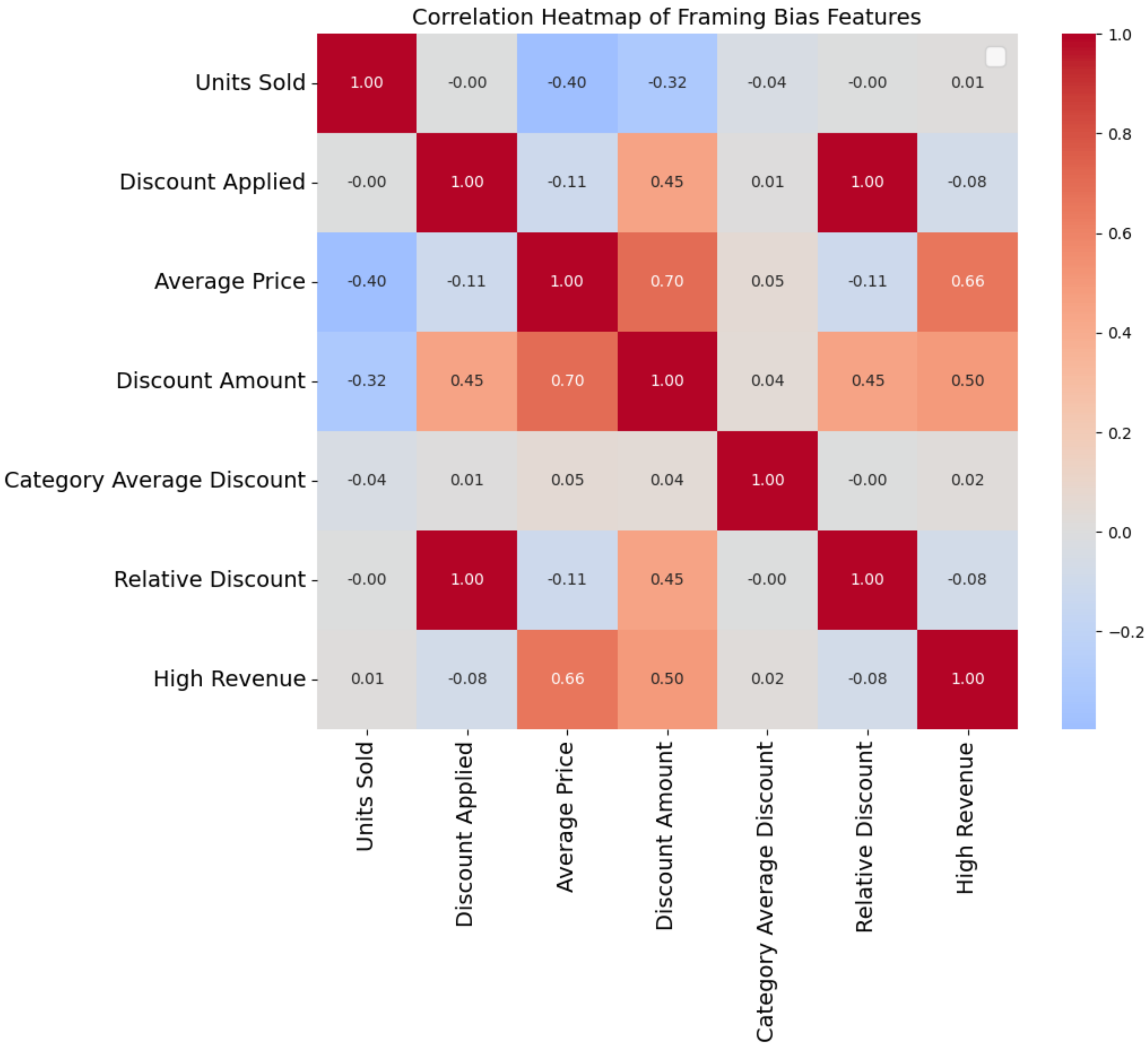


For social proof engineered features, the target variable (Recommended IND) forms relatively weak relationships with the other features. The non-target features formed very strong correlations with each other such as Review Count, Total Helpfulness and Social Proof Score, but their direct association to Recommended IND appears limited. Hence, while these features reflect peer influence, their explanatory strength for binary recommendation decision may be subtler.

Figure 16.
Correlation Heatmap of Social Proof Bias Features

Finally, in the framing dataset, the target variable (High Revenue Label) correlates moderately with Average Price and Discount Amount but forms a weaker association to Discount Applied and Relative Discount which are strongly interrelated, confirming earlier suspicion of multicollinearity. This demonstrates how discounting and pricing tactics interact, albeit no single variable entirely accounts for High Revenue classification.

Figure 17.
Correlation Heatmap of Framing Bias Features



Overall, the heatmaps showcase a descriptive summary of feature interactions within each dataset, offering insight into their influence on purchase decisions but avoiding any causal inference at this stage since later predictive modelling could indicate different findings.

6.4 Model Building and Training

Every model in this report was selected systematically, guided by existing research and established best practices in consumer behaviour prediction (Chapter 2). The selected algorithms are popular and well-documented in the field, guaranteeing their applicability and consistency with previous research. Each subsection contains a summary table of parameter and hyperparameter tuning choices, justifying its choice in maximising predictive power of each model.

6.4.1 Logistic Regression

This technique was selected for its interpretability, simplicity and long-standing work in consumer behaviour research (Van den Poel & Buckinx, 2005). It utilises a logistic function that constraints predictions between 0 and 1 to approximate the probability of an outcome, in this instance, a purchase. Therefore, coefficients can be connected to framing effects, social influences and urgency signals. Transparency like this is useful for confirming that results are consistent with behavioural theory. It is less effective at capturing how non-linear attributes interact, but it offers a reliable benchmark for evaluating the complexity of ensemble and neural models.

Table 12.
Parameter Summary Table for Logistic Regression

Aspect	Setting	Justification
Maximum Iterations	1000	Ensures convergence (increased from 100)
Penalty	L2 Regularisation	Shrinking large coefficients to reduce change of overfitting – important where there is correlation between engineered features
Regularisation Strength	1.0	Balances bias-variance trade-off. 1 is the default – not tuned.
Solver	lbfgs	Also the default. Good for larger datasets and handles L2 regularisation.
Random State	42	For reproducible data splitting for cross-model comparisons.

6.4.2 SVM

Linear Support Vector Machine (SVM) finds an optimal decision boundary that maximises the margin between data points from different classes, this allows SVM to outperform models like logistic regression when features are very separable. In consumer behaviour contexts with high-dimensional features, SVM has proven to be very robust, especially when using engineered predictors representing psychological qualities (Noti et al. 2016). Model transparency is less compared to logistic regression, but the linear kernel retains some to an extent through feature weights. Overall, this offers a compromise between interpretability and prediction power, serving as a bridge between simple linear models and more intricate ensemble approaches.

Table 13.
Parameter summary table for Linear SVM

Aspect	Setting	Justification
Maximum Iterations	1000	Ensures convergence
Penalty	L2 Regularisation	Shrinking large coefficients to reduce change of overfitting – important where there is correlation between engineered features

Regularisation Strength	1.0	Balances bias-variance trade-off. 1 is the default – not tuned.
Random State	42	For reproducible data splitting for cross-model comparisons.
Loss Function	Squared Hinge	Default for Linear SVM

6.4.3 Random Forest

Random Forest was selected as an ensemble method given its capacity to represent complex feature interactions and non-linear connections such as behavioural bias data (Balyemah et al., 2024). By being an ensemble of decision trees, it offers built-in measurements of feature importance, and reduced variance with bootstrapping (bagging). This improves insight into which engineered features influence decisions and increases predictive performance over non-ensemble techniques. In consumer purchase behaviour research, Random Forest has been successful (Chen et al., 2023) where the connection between transactional and psychological aspects is important. It offers effectiveness against overfitting deeming it even more suitable for this project.

Table 14.
Parameter and tuning summary table for Random Forest

Aspect	Setting	Justification
Number of estimators	200	More trees improve generalisation; balances training time and predictive power.
Maximum depth	None	Trees can grow fully, capturing complex bias effects.
Minimum Sample Split	2	Default settings with RFC – allows flexible learning from behavioural patterns.
Minimum Sample Leaf	1	
Bootstrap	True	Improves generalisation by ensuring diversity across trees.

6.4.4 XGBoost

Extreme Gradient Boosting (XGBoost) involves creating decision trees sequentially while fixing the mistakes of previous trees. Using this boosting approach when modelling more subtle, non-linear patterns such as the combined effects of social proof, pricing and urgency signals increases model effectiveness – consistently outperforming conventional models (Wang et al., 2023) - but also hinders model interpretability. Nevertheless, SHAP withstands this, allowing the behavioural importance of predictions to be recovered.

Table 15.
Parameter and tuning summary table for XGBoost

Aspect	Setting	Justification
Number of estimators	300	Ensures convergence, helps prevent excessive training time.
Maximum Depth	6	Balances model complexity and interpretability.
Learning Rate	0.05	Lower than default (0.1) – reduces overfitting while maintaining accuracy

Evaluation	Logloss	Appropriate for binary classification – directly boosts probabilistic predictions
Label Encoder	False	Disabled to make the model compatible with the scikit-learn pipelines.
Penalty	L2 Regularisation	Shrinking large coefficients to reduce change of overfitting – important where there is correlation between engineered features

6.4.5 Simple Neural Network

Implementing a simple neural network demonstrates the capacity for deep learning to model behavioural bias data without overcomplicating the model architecture, see Figure 18. Traditional models may overlook complex interactions between urgency and framing signals, for instance, but neural networks can convey this (Lang & Rettenmeier, 2017). When operating on larger datasets the model maintains its computational efficiency and prevents overfitting by using fewer dense layers, justifying its application in this report – acting as a stepping stone in determining if more in-depth evaluation and optimisation yield a better classification performance.

Figure 18.
Model architecture of simple neural network

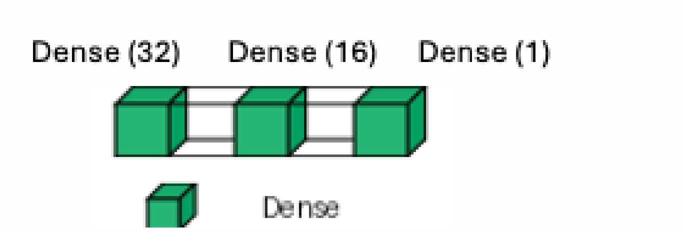


Table 16.
Parameter and Tuning Summary for a Simple Neural Network

Aspect	Setting	Justification
Model Architecture	Dense (32)	First hidden layer: 32 filters/kernels learn to recognise a different feature to form patterns.
	Dense (16)	Second hidden layer: 16 filters/kernels refine the feature representations the first layer learnt.
	Dense (1)	Output layer for binary classification, calculating the probability of belonging to a class.
Activation Function	Relu	Speeds up training in hidden layers
	Sigmoid	For binary classification
Optimiser	Adam	Adaptive learning rate, suited for engineered features
Loss function	Binary Crossentropy	Standard setting for binary classification tasks
Epochs	20	Maintain predictive power while preventing overfitting and minimising runtime.
Batch Size	32	

6.4.6 Enhanced Neural Network

Including batch normalisation, dropout regularisation and an increasing number of neurons in an enhanced neural network is expected to improve on a simple design (Figure 19). Such modifications aim to control overfitting while maximising the network's capacity to learn complex behavioural connections. Deep learning models, particularly more intricate ones, have proven to successfully identify patterns in high-dimensional consumer datasets, especially where psychological bias interactions are complex and non-linear (Vallarino, 2023). This method is the most sophisticated compared to the other methods, which aims to ascertain if increased model complexity results in the most substantial improvements of prediction power.

Figure 19.
Model architecture of enhanced neural network

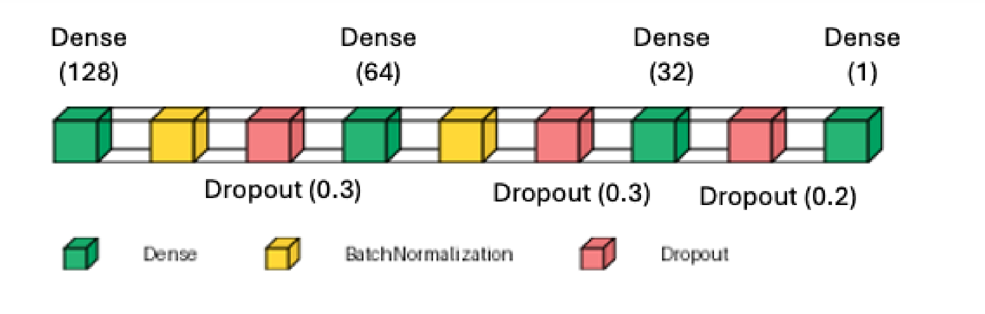


Table 17.
Parameter and tuning summary for enhanced neural network

Aspect	Setting	Justification
Model Architecture	Dense (128)	Larger first layer to capture even more complex feature interactions – more computationally expensive (counteracted by batch normalisation)
	Batch Normalisation	Stabilise and speeds up training by normalising activations across the batch.
	Dropout (0.3)	Reduces overfitting by randomly dropping neurons by 30%.
	Dense (64)	Reduced size to concentrate on learned information.
	Batch Normalization	Applied after second layer again to speed up process.
	Dropout(0.3)	Regularisation again
	Dense (32)	Reduced again to further focus on learned features
	Dropout (0.2)	Lowered to retain more features before output
	Dense (1)	Output layer for binary classification
Optimiser	Adam	Adaptive learning rate
Loss function	Binary Crossentropy	For binary classification
Epochs	30	By increasing the number of epochs, it leads to smoother but less frequent learning adjustments

6.5 Evaluation Metrics

Model evaluation was conducted across four dimensions: performance overview, generalisation, error analysis and interpretability. Table 18 summarises the evaluation framework adopted across all models.

Table 18.
Evaluation Metrics Overview

Aspect	Description
Performance Overview	
Classification Report	Measures precision, recall, F1-score and accuracy to assess predictive power of models
Training time	Monitors efficiency of each model but recording its algorithm execution time
Generalisation Check	
Overfitting Gap	$Overfit/Underfit\ Gap = Validation\ Loss\ (last\ epoch) - Training\ Loss\ (last\ epoch)$ The overfit gap is: • Likely underfitting if < 0 • Good generalisation if ≈ 0 • Likely overfitting if > 0
Error Analysis	
Confusion Matrix	Full breakdown of correct and incorrect predictions across both classes, highlighting the number of true positives/negatives and false positives/negatives.
Interpretability	
SHAP Analysis	Global summary plots used to demonstrate feature importance and direction of effect.

6.6 Design Trade-Offs

Throughout the entire research design process, some trade-offs were flagged and deemed necessary to balance the rigour of the methodology, interpretability and feasibility. Table 19 summarises these trade-offs.

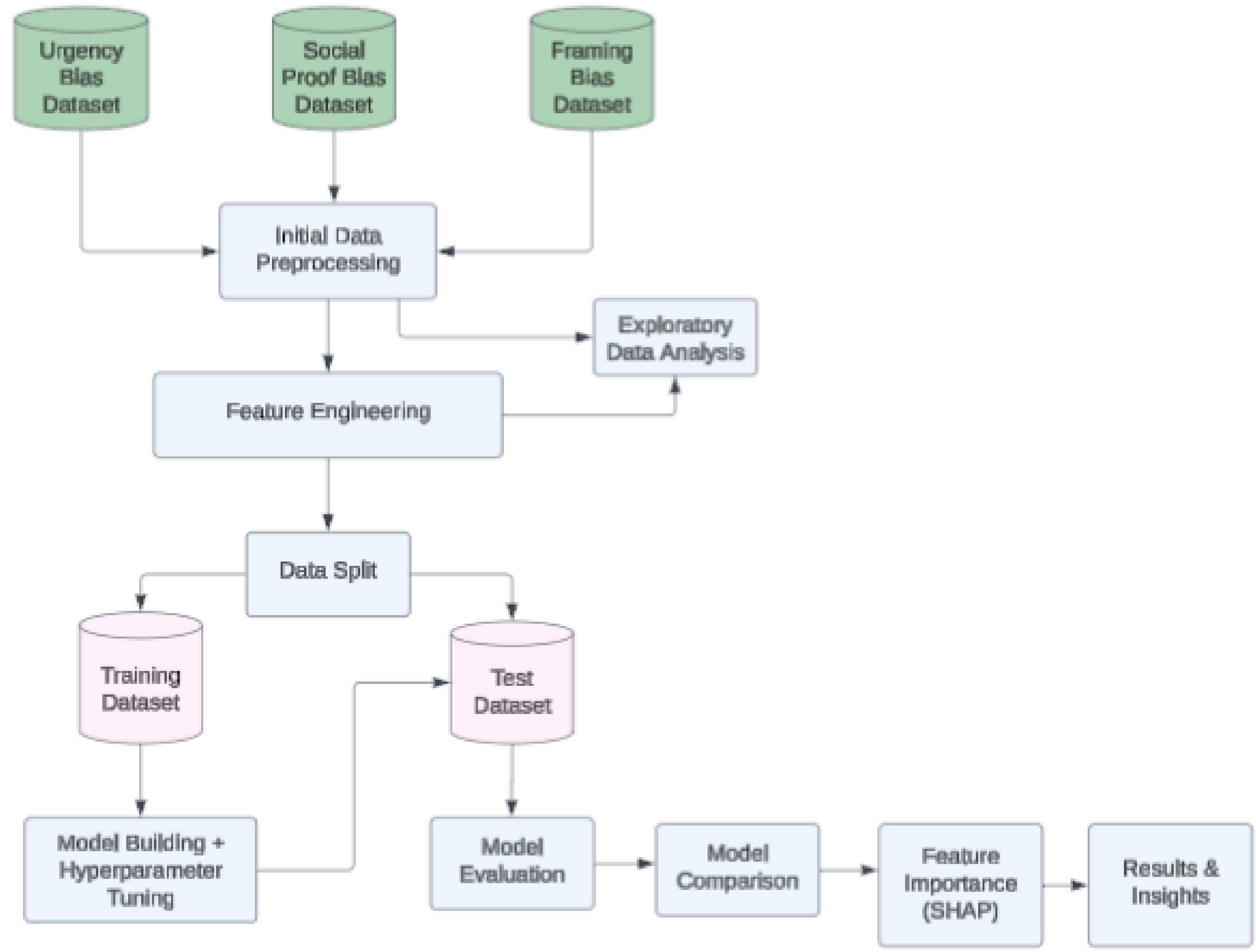
Table 19.
Design Trade-Offs in Methodology

Design Decision	Rationale	Trade-Off
Choice of Dataset	To model psychological biases, this study used openly accessible e-commerce datasets.	The datasets are readily available and large, but they are noisy, imbalanced and there is little control over data quality.
Feature Engineering	Converted urgency, social proof and framing biases into predictor cues to predict consumer purchase likelihood.	Operationalises the abstract biases but at risk of oversimplifying human behaviour.
Model Selection	Included baseline, interpretable models (Logistic Regression, SVM) and complex models (Random Forest, XGBoost and Neural Networks)	Baseline models are simpler and more transparent while increasing the complexity of models usually improves accuracy but becomes less transparent.

6.7 Chapter Summary

The flowchart below summarises the methodology process.

Figure 20.
Flowchart summarising the methodology



Chapter 7:

Results and Evaluation

7.1 Urgency Bias Results

7.1.1 Classification Report

Table 20.
Classification Report Across All Models for Urgency Bias

	Logistic Regression	Linear SVM	XGBoost	Random Forest	Simple Neural Network	Enhanced Neural Network
Precision (not urgent)	0.91	0.91	0.91	0.92	0.89	0.91
Precision (urgent)	1.00	1.00	0.97	0.79	0.99	1.00
Recall (not urgent)	1.00	1.00	0.99	0.96	0.99	1.00
Recall (urgent)	0.59	0.59	0.60	0.68	0.53	0.58
F1-Score (not urgent)	0.95	0.95	0.95	0.94	0.95	0.95
F1-Score (urgent)	0.74	0.74	0.74	0.73	0.69	0.74
Accuracy	0.92	0.92	0.92	0.95	0.91	0.92

All models performed strongly on the urgency bias dataset, with accuracy scores ranging from 91% to 95%, although, class-level analysis uncovers important differences in each model's ability to handle urgent versus non-urgent predictions.

Performance was almost flawless across models for the non-urgent instances, exceeding 0.89 in precision, recall and F1-scores. In particular, the enhanced neural network and logistic regression models attained a perfect recall and 0.91 precision, implying extremely low false negatives.

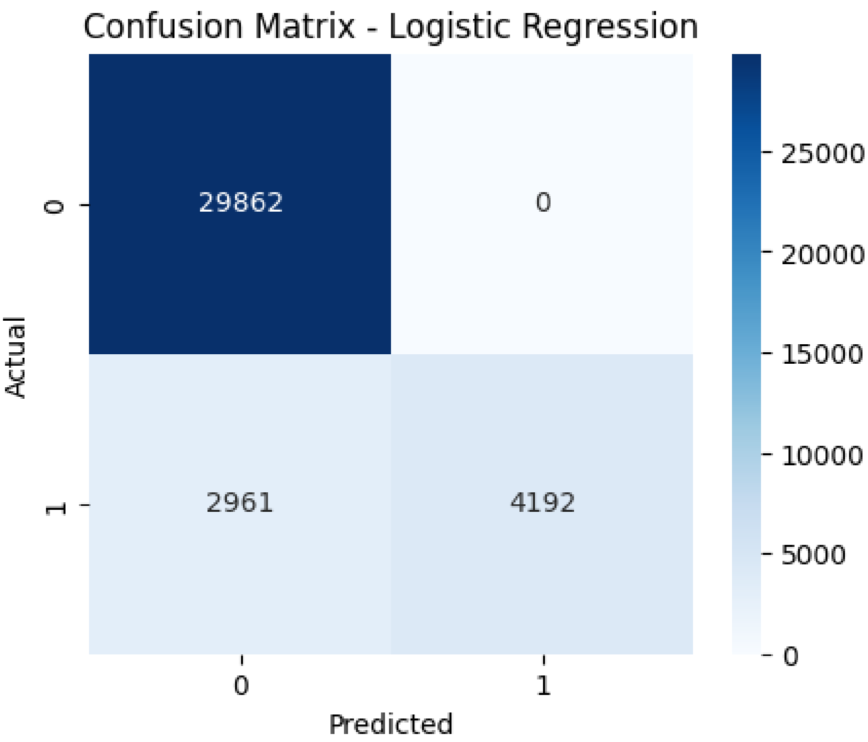
However, predicting 'urgent' cases proved more difficult. Various models such as Logistic Regression, Linear SVM, XGBoost and Enhanced Neural Network all had relatively high F1-scores of 0.74, though this was subsequent with different trade-offs. For instance, Logistic Regression, Linear SVM and the Enhanced Neural Network had perfect precision, with moderate recall (0.59), demonstrating many false negatives despite precise predictions when urgency was flagged. Whereas, with Random Forest, recall increased to 0.68 but sacrificed precision (0.79), implying a more sensitive model but with greater false positives. Similarly, the Simple Neural Network showed a tendency to ignore urgent instances, with the lowest F1-score (0.69) and recall (0.53).

Given that all models yielded very similar results, they would all be suitable for detecting urgeny bias-induced purchases. Although, XGBoost and Random Forest provided the greatest balance.

7.1.2 Confusion Matrix

Figure 21.

(a) Confusion Matrix for Logistic Regression

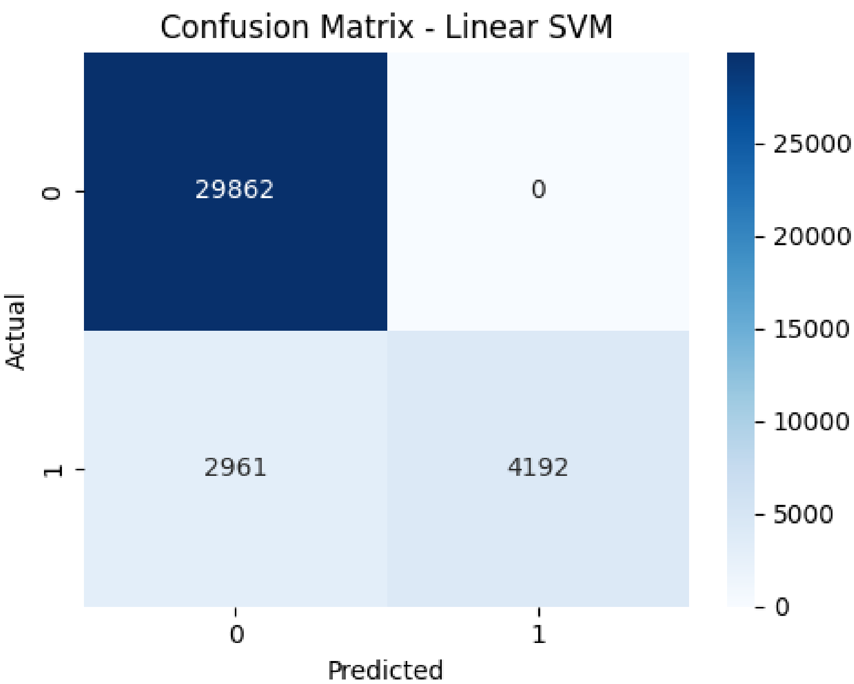


(b) Confusion Matrix Interpretation

Number of Instances	37,015
True Negative	29,862
False Negative	2,961
False Positive	0
True Positive	4,192
Total Correct Predictions	34,054
% of Correct Prediction	92.00%
Total Incorrect Predictions	2,961
% of Incorrect Predictions	8.00%

Figure 22.

(a) Confusion Matrix for SVM



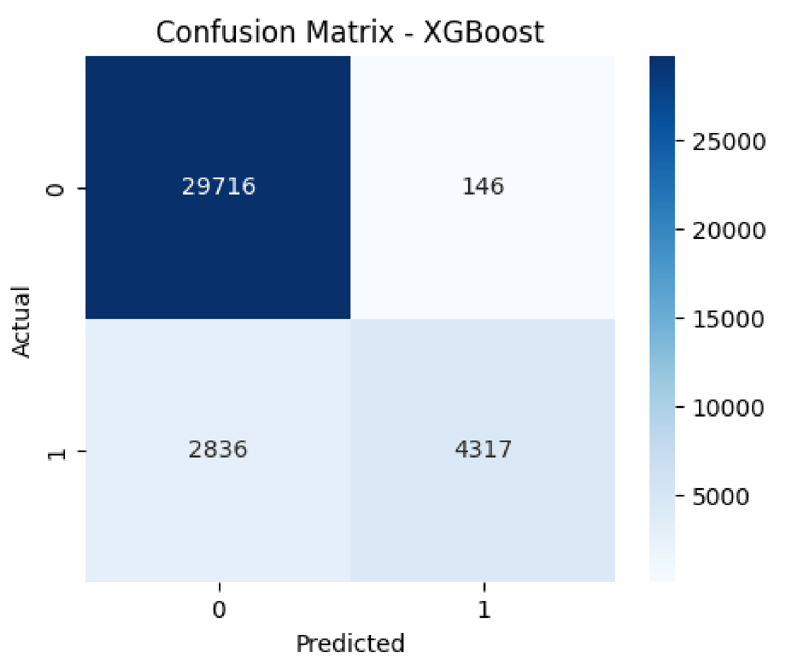
(b) Confusion Matrix Interpretation

Number of Instances	37,015
True Negative	29,862
False Negative	2,961
False Positive	0
True Positive	4,192
Total Correct Predictions	34,054
% of Correct Prediction	92.00%
Total Incorrect Predictions	2,961
% of Incorrect Predictions	8.00%

Both linear models achieved an identical confusion matrix with 92% accuracy, accurately predicting all positive instances (no false positives), and many negative cases. Although it overlooked many urgent cases (2,961 false negatives) – further demonstrating high precision but low recall for predicting urgent instances.

Figure 23.

(a) Confusion Matrix for XGBoost



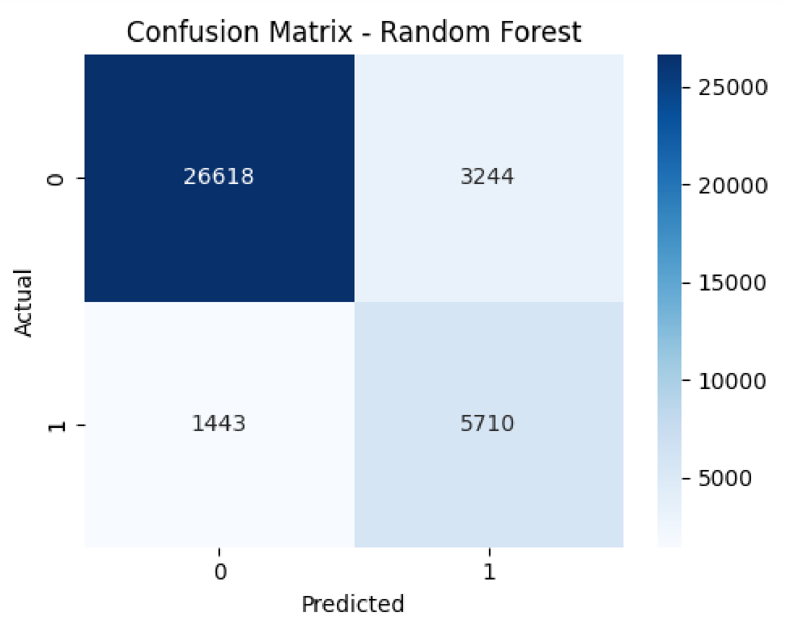
(b) Confusion Matrix Interpretation

Number of Instances	37,015
True Negative	29,716
False Negative	2,836
False Positive	146
True Positive	4,317
Total Correct Predictions	34,033
% of Correct Prediction	91.94%
Total Incorrect Predictions	2,982
% of Incorrect Predictions	8.06%

XGBoost accurately predicted 91.94% of all cases. It excelled in both classes but with 2.836 false negatives and 146 false positives, showing better detection of non-urgent cases than urgent ones.

Figure 24.

(a) Confusion Matrix for Random Forest



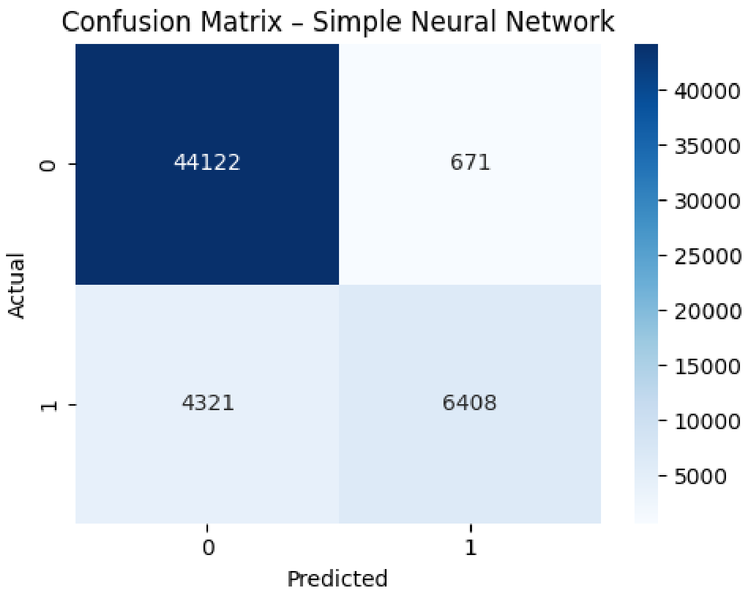
(b) Confusion Matrix Interpretation

Number of Instances	37,015
True Negative	26,618
False Negative	1,443
False Positive	3,244
True Positive	5,710
Total Correct Predictions	32,328
% of Correct Predictions	87.34%
Total Incorrect Predictions	4,687
% of Incorrect Predictions	12,66%

Random Forest was the lowest performing model with 87.34% instances correctly predicted – a high score nonetheless, but low relative to other models. With a higher rate of false positives (3,244), it sometimes misclassified non-urgent classes as urgent. False negatives were relatively lower at 1,443.

Figure 25.

(a) Confusion Matrix for Simple Neural Network



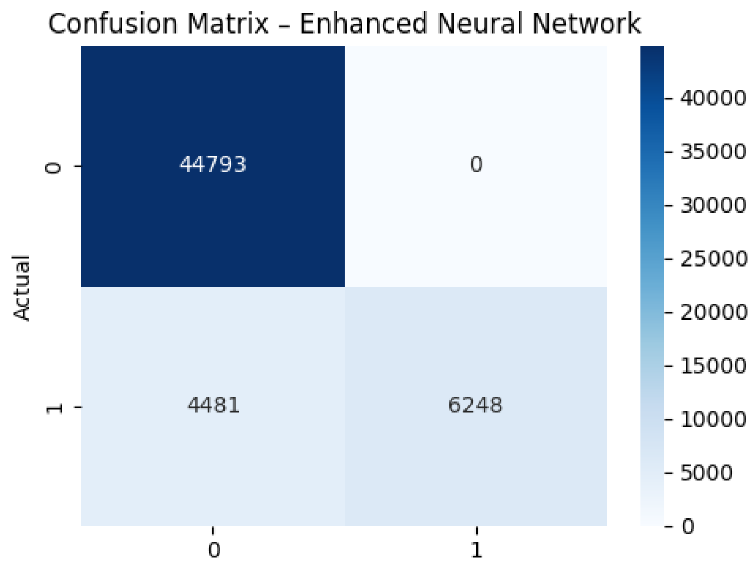
(b) Confusion Matrix Interpretation

Number of Instances	55,522
True Negative	44,122
False Negative	4,321
False Positive	671
True Positive	6,408
Total Correct Predictions	50,530
% of Correct Prediction	91.30%
Total Incorrect Predictions	4,992
% of Incorrect Predictions	8.70%

This model resulted in strong performance across both classes as well, with 91.30% accuracy. It demonstrated reliable urgent class detection with a low false positive rate (671), although there is still room for improvement for the recall of urgent predictions given the 4,321 false negatives.

Figure 26.

(a) Confusion Matrix for Enhanced Neural Network



(b) Confusion Matrix Interpretation

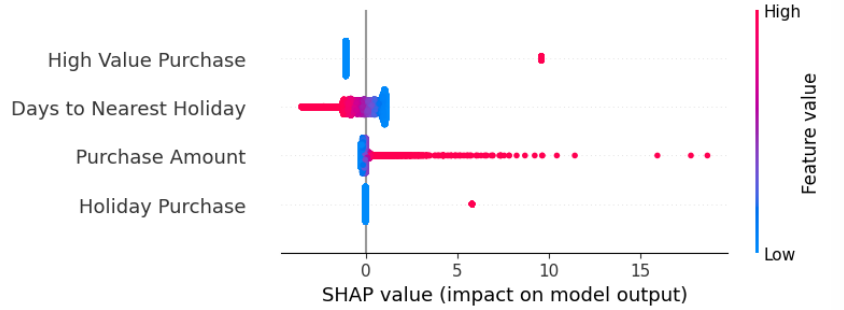
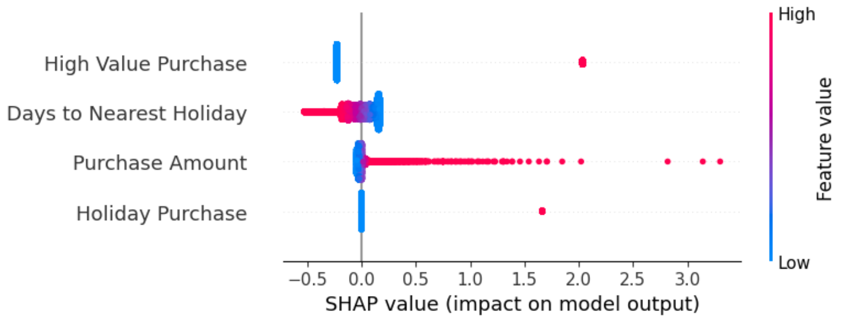
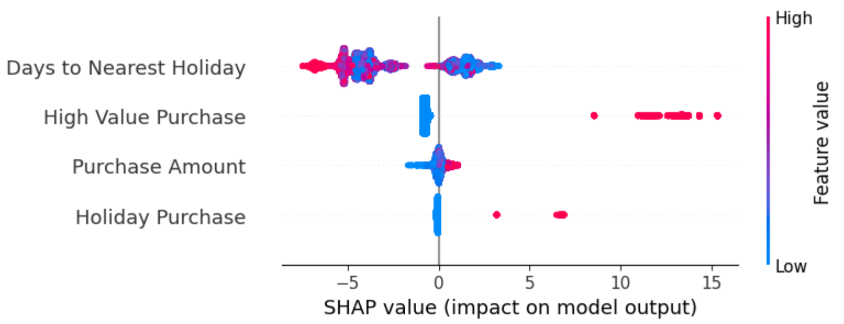
Number of Instances	55,522
True Negative	41,793
False Negative	4,481
False Positive	0
True Positive	6,248
Total Correct Predictions	48,041
% of Correct Prediction	91.67%
Total Incorrect Predictions	4,481
% of Incorrect Predictions	8.33%

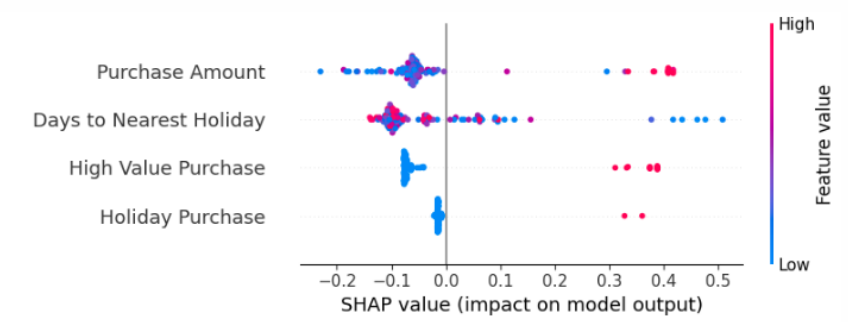
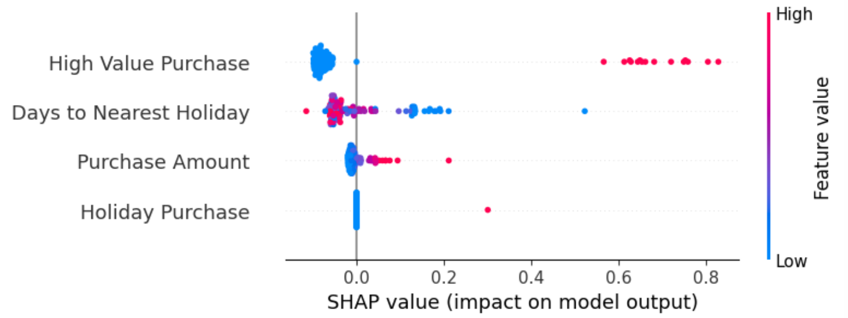
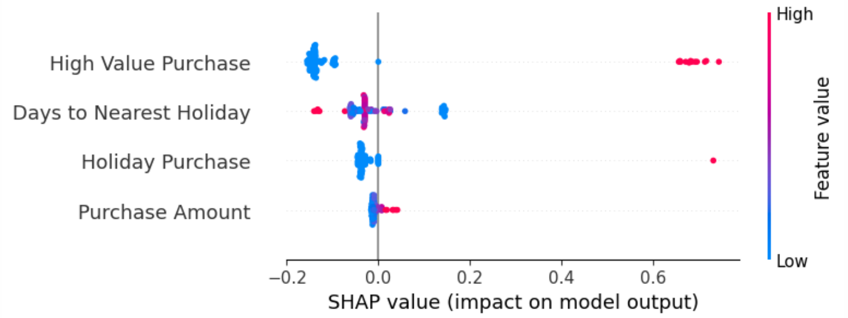
Like Logistic Regression and Linear SVM, this model also achieved zero false positives, proving to be highly reliable. Although, there is still room for improvement in improving recall while preserving a high precision given that 4,481 false negatives indicate urgent instances were missed.

7.1.3 Further Evaluation Metrics + SHAP for Urgency Bias

Table 21.

Table comparing training time, generalisation (overfit gap) and most influential features in urgency bias-induced purchases across models

<table>
<tr><th>Model</th><th>Training Time (s)</th><th>Overfit Gap</th><th>Top Feature</th></tr>
<tr><td>Logistic Regression</td><td>1.402</td><td>0.003</td><td>
</td></tr>
<tr><td>Linear SVM</td><td>13.091</td><td>0.003</td><td>
</td></tr>
<tr><td>XGBoost</td><td>0.887</td><td>0.000</td><td>
</td></tr>
</table>

Random Forest	12.09	0.049	
Simple Neural Network	407.03	0.003	
Enhanced Neural Network	554.048	0.000	

XGBoost appears to be the fastest performing model (0.89s) with no overfitting, making it excellent for high-performance use in real-world situations. Logistic Regression and Linear SVM are also quick (1.4s and 13s, respectively) with minimal overfitting, which offers a strong baseline performance with good generalisation. Both Neural Networks took significantly longer to train, which limits their practicality in real-time applications despite a high accuracy. Although, the Enhanced Neural Network showed perfect generalisation with no overfitting. While Random Forest was effective, it has the highest overfit gap (0.049), showing potential concerns with a lack of generalisation or complexity.

SHAP analysis consistently identified High Value Purchase and Days to Nearest Holiday as the most influential factors overall. All models revealed this as the top features, except for Random Forest which was most influenced by Purchase Amount. This validates the feature engineering conducted, emphasising the psychological motivator in urgency driven purchases during urgency-induced situations.

Overall, all models excelled. Random Forest was still well-performing but was moderately outperformed by the remaining five models, deeming it the least suitable for detecting urgency bias. XGBoost can be considered the best-performing model given its balance in fast processing, perfect generalisation, high accuracy and precision, recall and f1-score. The SHAP results highlights that temporal context and value considerably influence urgency and decision-making, which supports theoretical expectations.

7.2 Social Proof

7.2.1 Classification Report

Table 22.
Classification Report Across All Models for Social Proof

	Logistic Regression	**Linear SVM**	**XGBoost**	**Random Forest**	**Simple Neural Network**	**Enhanced Neural Network**
Precision (not recommended)	0.68	0.78	0.51	0.51	0.78	0.73
Precision (recommended)	0.83	0.83	0.83	0.83	0.83	0.83
Recall (not recommended)	0.06	0.05	0.04	0.04	0.05	0.05
Recall (recommended)	0.99	0.99	0.99	0.99	0.99	0.99
F1-Score (not recommended)	0.11	0.09	0.07	0.07	0.09	0.10
F1-Score (recommended)	0.90	0.91	0.90	0.90	0.91	0.91
Accuracy	0.83	0.83	0.82	0.82	0.83	0.83

Performance in the recommended class is consistently high across all models. Precision scores stay consistent at 0.83 but recall for all models is near-perfect (0.99), resulting in high F1-scores ranging from 0.90-0.91. Therefore, these models were extremely sensitive to socially influenced purchases, hardly failing to classify them as recommended. However, classifying the not-recommended class was generally poor. All models commonly misclassified these purchases as recommended, as suggested by low recall values (0.04-0.06). Despite variation in precision (0.51-0.78), F1-score remained very low, below 0.11, for this class with reveals a significant imbalance between precision and recall, regardless of model architecture.

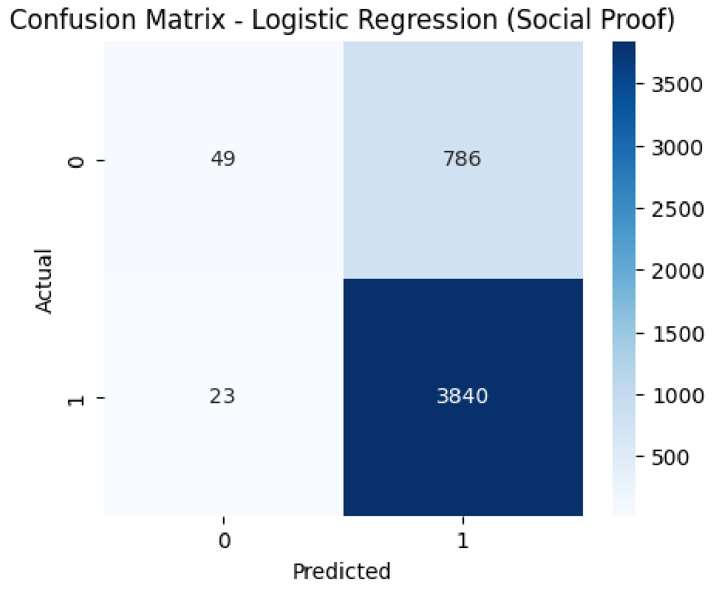
Even though accuracy is consistent across all models, it makes it difficult to judge performance. Linear SVM, Enhanced Neural Network and Simple Neural Network only marginally outperform other models in F1-score for recommended instance, but there is not a clear superior model with balance across both classes.

All models are reliable in identifying social proof-influenced purchases, but none significantly outperform others. Performance differences across all metrics are minimal, and model choice would be more reasonable and realistic by depending on interpretability, training time and generalisation than marginal metric differences.

7.2.2 Confusion matrix

Figure 27.

(a) Confusion Matrix for Logistic Regression



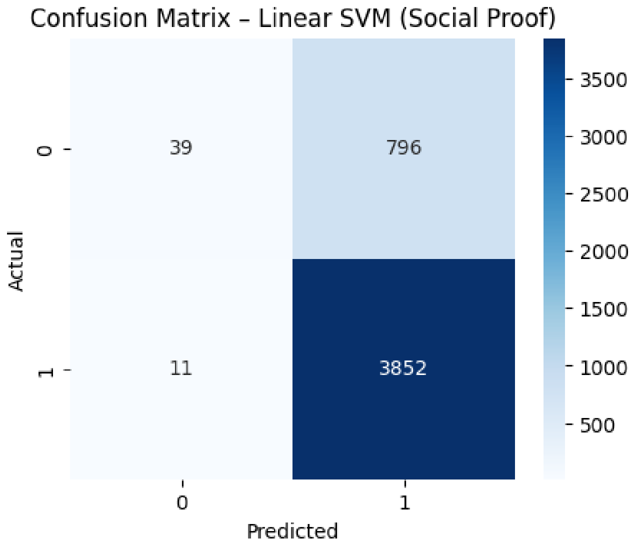
(b) Confusion Matrix Interpretation

Number of Instances	4,698
True Negative	49
False Negative	23
False Positive	786
True Positive	3,840
Total Correct Predictions	3,889
% of Correct Prediction	82.76%
Total Incorrect Predictions	809
% of Incorrect Predictions	17.24%

Logistic Regression correctly classified 3,889 cases out of 4,698 instances, achieving 82.76% accuracy. 786 false positives do further indicate over-prediction of the recommended class, while only 49 true negatives further support difficulty in classifying non-recommended instances

Figure 28.

(a) Confusion Matrix for SVM



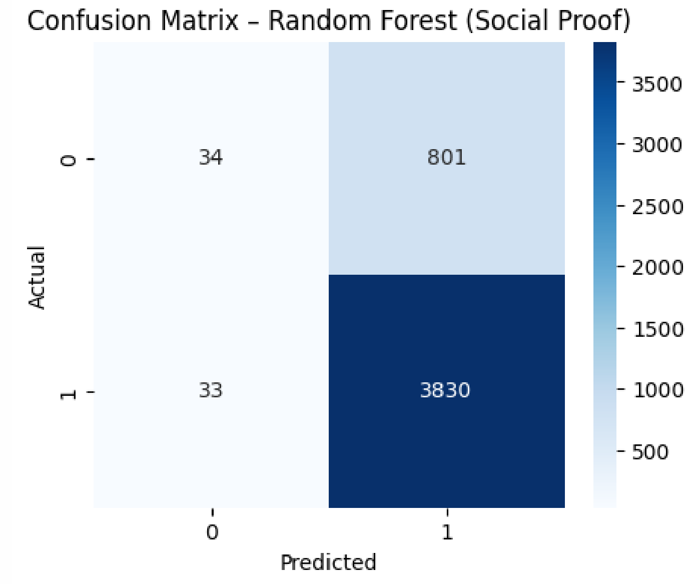
(b) Confusion Matrix Interpretation

Number of Instances	4,698
True Negative	39
False Negative	11
False Positive	796
True Positive	3,852
Total Correct Predictions	3,891
% of Correct Prediction	82.80%
Total Incorrect Predictions	807
% of Incorrect Predictions	17.20%

This model demonstrates strong performance overall with 82.80% accuracy, although 796 false positives compared to only 39 true negatives imply poor classification of the non-recommended class, despite well classified recommended instances.

Figure 29.

(a) Confusion Matrix for Random Forest



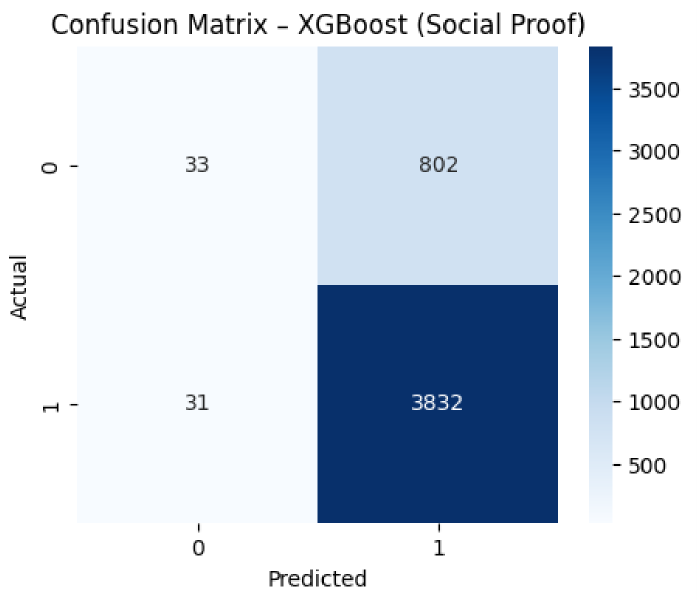
(b) Confusion Matrix Interpretation

Number of Instances	4,698
True Negative	34
False Negative	23
False Positive	801
True Positive	3,830
Total Correct Predictions	3,864
% of Correct Predictions	82.45%
Total Incorrect Predictions	824
% of Incorrect Predictions	17.55%

With 3,830 true positives, this model achieves 82.45% accuracy. Although, again, it struggles to identify the negative (non-recommended) class, given the 801 false positives and only 34 true negatives, which reflects a possible class imbalance.

Figure 30.

(a) Confusion Matrix for XGBoost



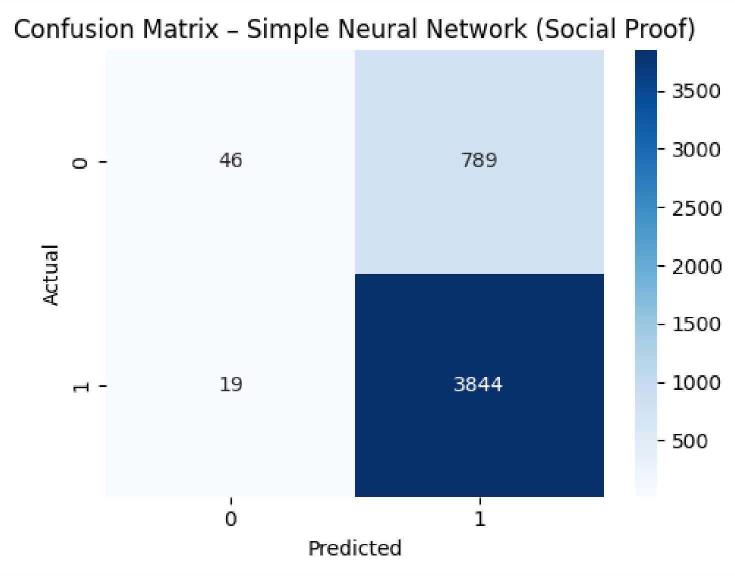
(b) Confusion Matrix Interpretation

Number of Instances	4,698
True Negative	33
False Negative	31
False Positive	802
True Positive	3,832
Total Correct Predictions	3,865
% of Correct Prediction	82.30%
Total Incorrect Predictions	833
% of Incorrect Predictions	17.70%

XGBoost correctly predicted most urgent cases (3,832 true positives), achieving 82.30% accuracy, although it shows a substantial bias towards the recommended class with only 33 true negatives and 802 false positives, making it less reliable for non-recommended purchase detection.

Figure 31.

(a) Confusion Matrix for Simple Neural Network



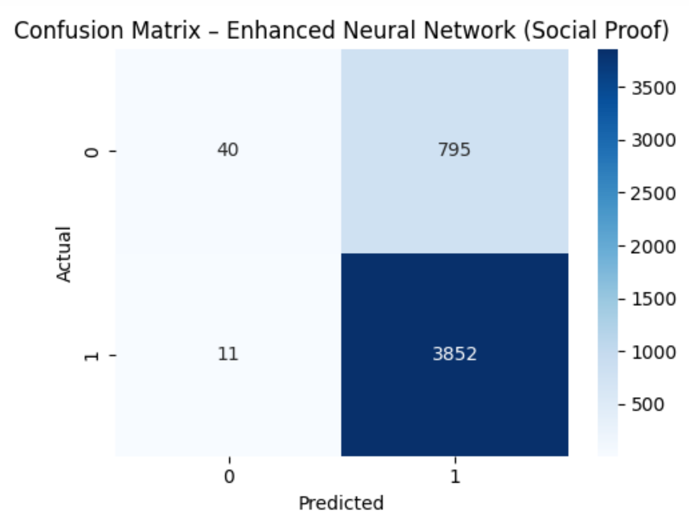
(b) Confusion Matrix Interpretation

Number of Instances	4,698
True Negative	46
False Negative	19
False Positive	789
True Positive	3,844
Total Correct Predictions	3,890
% of Correct Prediction	82.80%
Total Incorrect Predictions	808
% of Incorrect Predictions	17.20%

This model correctly identified 3,844 recommended instances and 46 non-recommended. Classification performance may be strong for recommended predictions, but with 789 false positives, it signifies a notable tendency to overpredict recommended purchases.

Figure 32.

(a) Confusion Matrix for Enhanced Neural Network



(b) Confusion Matrix Interpretation

Number of Instances	4,698
True Negative	40
False Negative	11
False Positive	795
True Positive	3,852
Total Correct Predictions	3,892
% of Correct Prediction	82.84%
Total Incorrect Predictions	806
% of Incorrect Predictions	17.16%

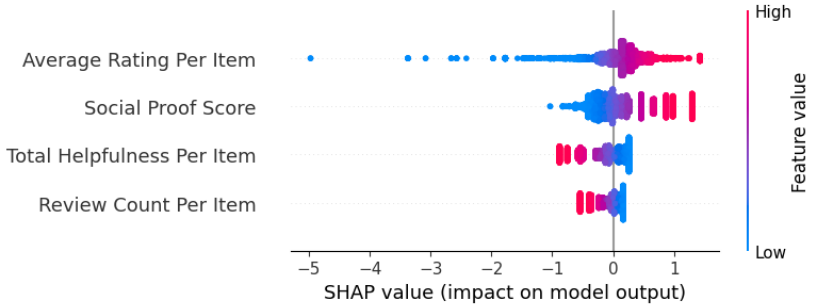
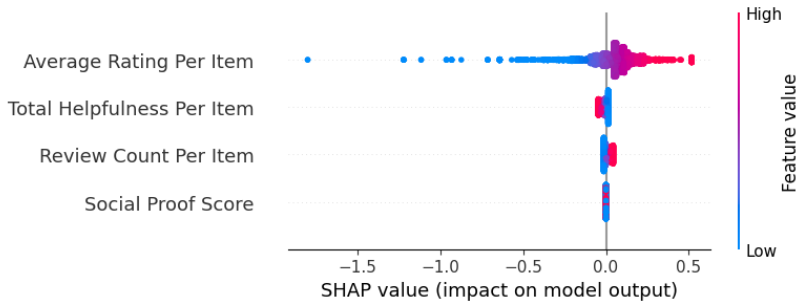
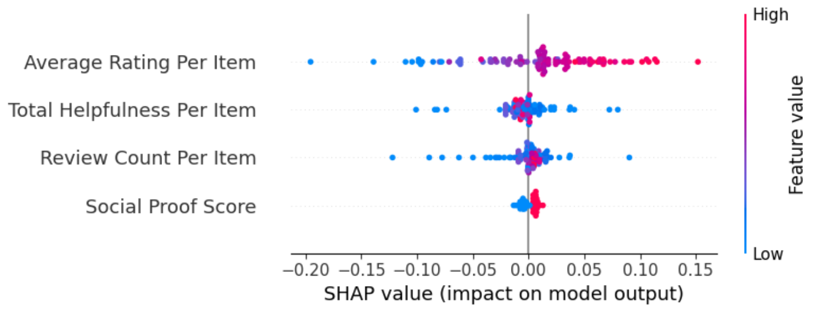
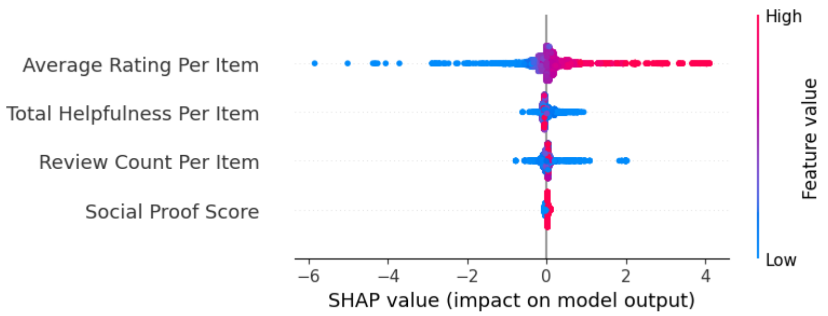
With 3,852 recommended and 40 non-recommended instances correctly predicted, the Enhanced Neural Network achieves 82.84% accuracy. Although, like the previous models, 795 false positives suggest a bias to overpredicting recommended cases.

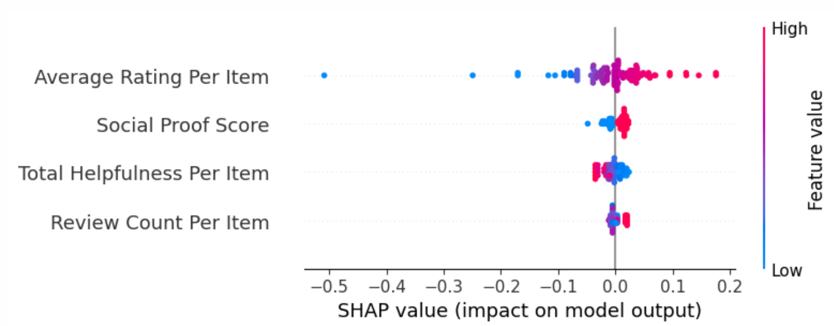
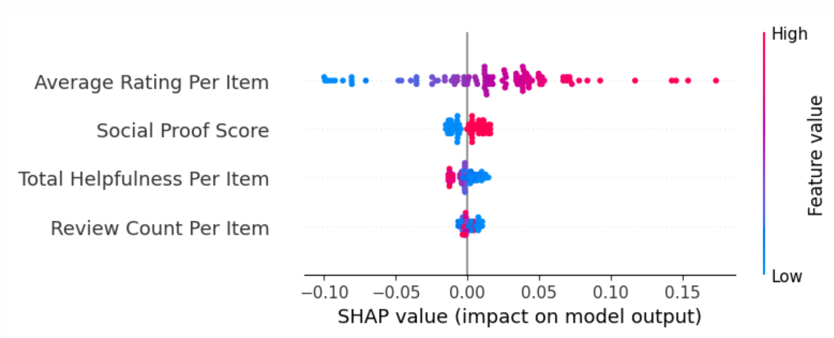
All confusion matrixes revealed similar findings, there is no clear superior or inferior model, hence further evaluation is required to identify this, see Table 23.

7.2.3 *Further Evaluation Metrics + SHAP for Social Proof Bias*

Table 23.

Table comparing training time, generalisation (overfit gap) and most influential features in social proof bias-induced purchases across models

Model	Training Time (s)	Overfit Gap	Top Feature
Logistic Regression	0.300	0.001	
Linear SVM	0.179	0.002	
Random Forest	3.176	0.006	
XGBoost	2.319	0.006	

<table>
<tr><td>Simple Neural Network</td><td>55.299</td><td>0.002</td><td>
</td></tr>
<tr><td>Enhanced Neural Network</td><td>68.051</td><td>0.000</td><td>
</td></tr>
</table>

As seen in the SHAP plots, 'Average Rating Per Item' and 'Total Helpfulness Per Item' were the overall strongest motivator of predicted outcomes. Instances were more likely to be classified as "recommended" if average ratings were higher, aligning with psychological theory that perceived quality strongly influences consumer decision-making. Likewise, endorsement from other users reinforced social proof given reviews marked as helpful, contributing to helpfulness per item scores, enhanced this effect. Tree-based models such as Random Forest and XGBoost weighed 'Review Count Per Item' more heavily, highlighting its ability to capture non-linear patterns such as between popularity and likelihood of recommendations. The engineered feature (social proof score) added theoretical value having combined quality, popularity and endorsement into a single feature. Although its SHAP importance ranked low in Linear SVM and the tree-based model, it ranked high in Logistic Regression and Neural Network Models, offering interpretability benefits by bridging the gap between behavioural phenomena and computational modelling.

XGBoost showed the most balanced performance with high accuracy, quick training time and a minimal overfit gap. Random Forest performed just as well with a slightly longer training time, and slightly more variance across predictions. Logistic Regression and SVM performed the fastest with negligible overfitting, although their feature weighting was nuanced, restricting its ability to use engineered social proof signals. Both Neural Networks performed with great stability and generalisation, although at the expense of a very long training period, albeit they have a very high predictive accuracy, but it was computationally wasteful.

Collectively, XGBoost and Random Forest most effectively capture social proof bias.

7.3 Framing

7.3.1 Classification Report

Table 24.
Classification Report Across All Models for Urgency Bias

	Logistic Regression	**Linear SVM**	**XGBoost**	**Random Forest**	**Simple Neural Network**	**Enhanced Neural Network**
Precision (not high revenue)	0.95	0.94	0.99	0.99	1.0	0.99
Precision (high revenue)	0.96	0.96	1.00	1.00	0.99	1.0
Recall (not high revenue)	0.96	0.96	0.99	0.99	0.99	1.0
Recall (high revenue)	0.95	0.94	0.99	0.99	1.0	0.99
F1-Score (not high revenue)	0.95	0.95	1.00	1.00	1.0	1.0
F1-Score (high revenue)	0.95	0.95	1.0	1.0	1.0	1.0
Accuracy	0.95	0.95	0.99	0.99	1.00	1.0

All models performed with great predictive power. Logistic Regression and SVM performed the worst relative to other models, but they still achieved great levels of precision, recall and F1-scores for both classes. Even basic linear models captured framing effects effectively.

The ensemble models improved in scores across all metrics, with near-perfect results in both classes – demonstrating that they could capture framing bias between classes without favouring either result.

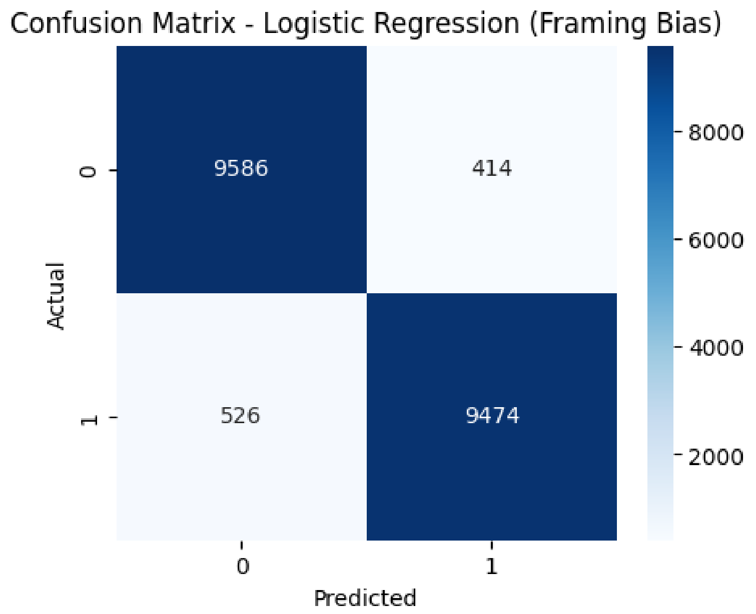
Neural networks somewhat performed better than the ensemble approaches, but the difference was minor. Both models achieved essentially perfect balanced performance with all metrics ranging from 0.99 to 1.

While there were minor differences in efficiency amongst models, all were extremely effective at predicting framing bias. The overall findings suggest framing was the easiest bias to classify with little variation between models.

7.3.2 *Confusion Matrix*

Figure 33.

(a) Confusion Matrix for Logistic Regression



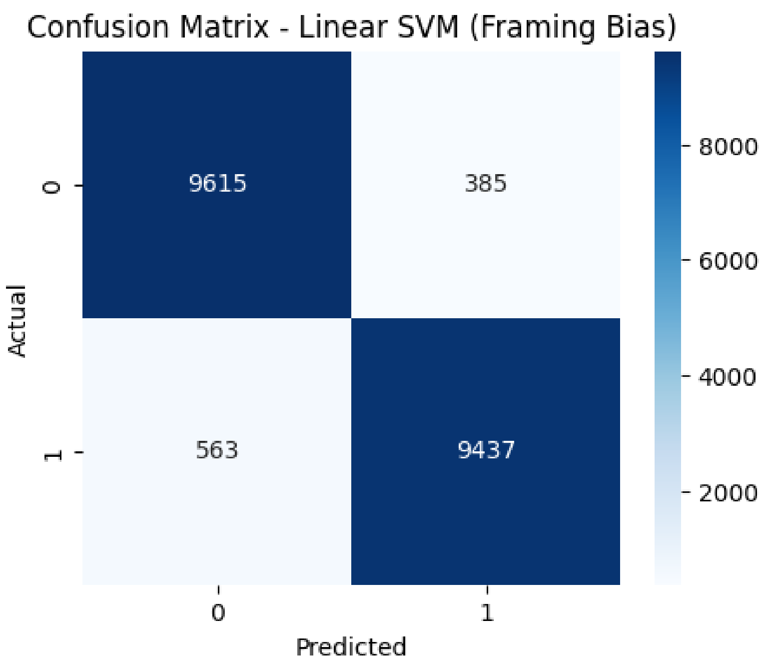
(b) Confusion Matrix Interpretation

Number of Instances	19,694
True Negative	9,580
False Negative	526
False Positive	414
True Positive	9,174
Total Correct Predictions	18,754
% of Correct Prediction	95.23%
Total Incorrect Predictions	940
% of Incorrect Predictions	4.77%

Logistic Regression made 18,754 correct predictions of out 19,694, resulting in 95.23% accuracy. There was minimal difference in false negatives and positives showing errors were minimal, indicating consistent and balanced classification across both 'high revenue' and 'not high revenue' classes

Figure 34.

(a) Confusion Matrix for SVM



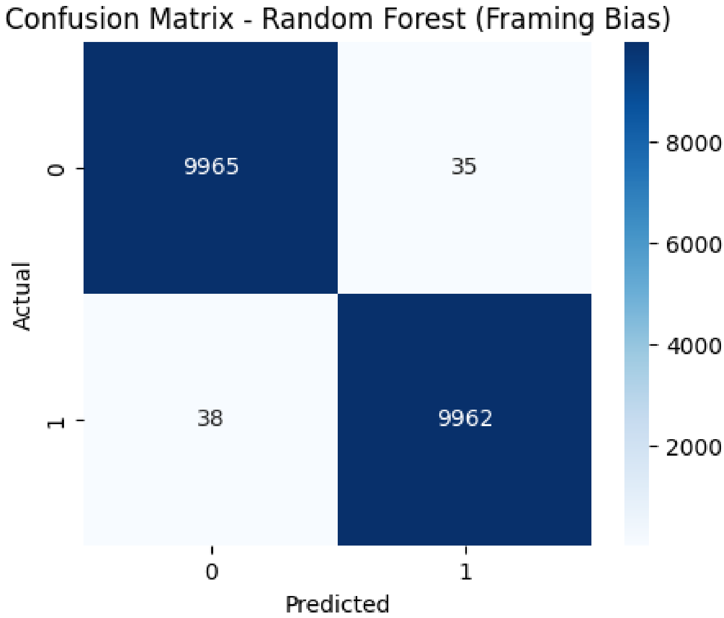
(b) Confusion Matrix Interpretation

Number of Instances	19,700
True Negative	9,615
False Negative	563
False Positive	385
True Positive	9,137
Total Correct Predictions	18,752
% of Correct Prediction	95.20%
Total Incorrect Predictions	948
% of Incorrect Predictions	4.80%

Linear SVM performed similarly to Logistic Regression, with slightly more errors given a greater difference between false negatives and positives. Nevertheless, the difference is still very minimal, so this model also offered a balanced and reliable performance across both classes.

Figure 35.

(a) Confusion Matrix for Random Forest



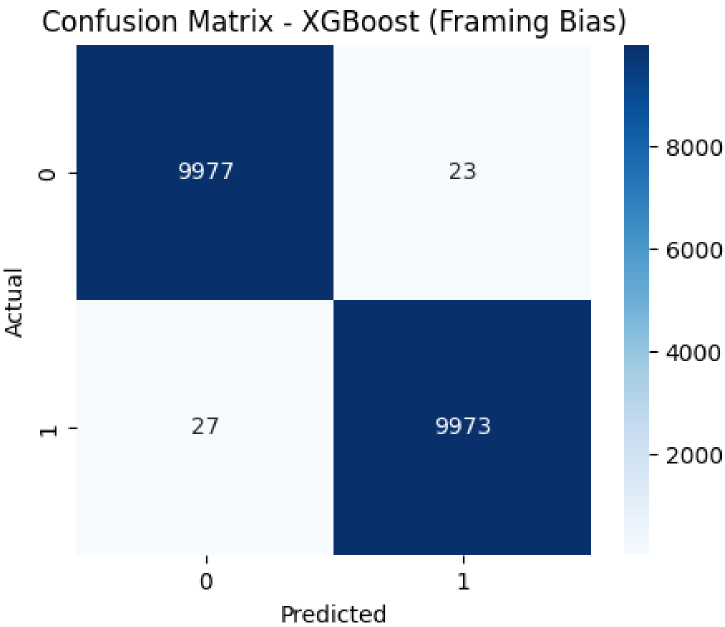
(b) Confusion Matrix Interpretation

Number of Instances	20,000
True Negative	9,965
False Negative	38
False Positive	35
True Positive	9,962
Total Correct Predictions	19,927
% of Correct Prediction	99.64%
Total Incorrect Predictions	73
% of Incorrect Predictions	0.36%

This model also excelled, although it has been slightly outperformed by XGBoost with 38 false negatives and 35 false positives – greater than that of XGBoost. Nevertheless, out of 20,000 instances, Random Forest correctly classified 19,927 of these, resulting in 99.64% accuracy.

Figure 36.

(a) Confusion Matrix for XGBoost



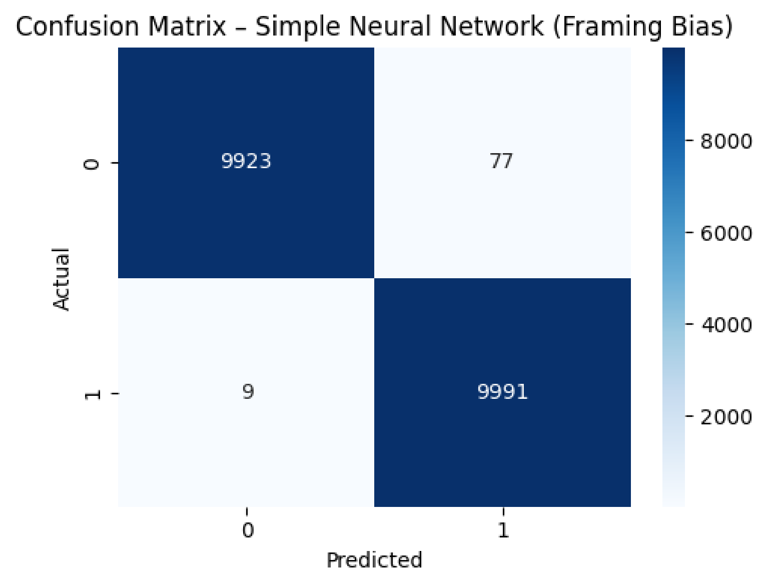
(b) Confusion Matrix Interpretation

Number of Instances	20,000
True Negative	9,977
False Negative	27
False Positive	23
True Positive	9,973
Total Correct Predictions	19,950
% of Correct Prediction	99.75%
Total Incorrect Predictions	50
% of Incorrect Predictions	0.25%

This confusion matrix reveals the distinct improvement in predictive accuracy from the linear models. XGBoost achieved 99.75% accuracy with 19,950 correct predictions out of 20,000 instances. Incorrect classifications were extremely rare given only 27 false negatives and 23 false positives, conveying near-perfect classification performance of both classes.

Figure 37.

(a) Confusion Matrix for Simple Neural Network



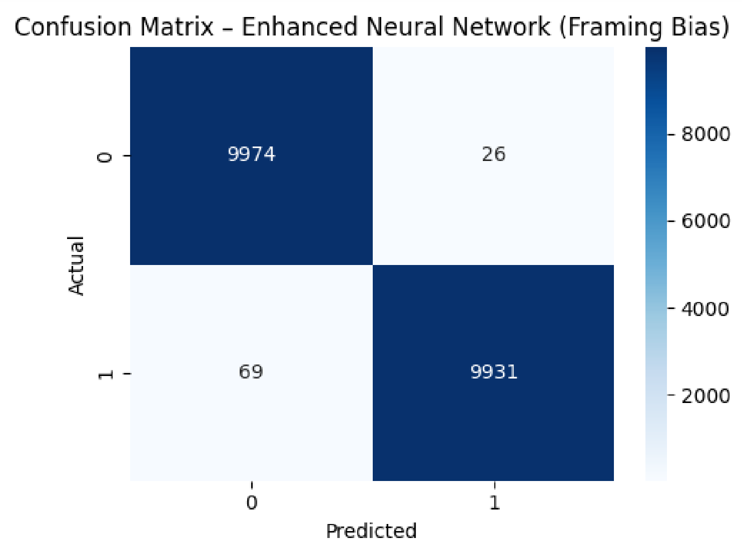
(b) Confusion Matrix Interpretation

Number of Instances	20,000
True Negative	9,923
False Negative	9
False Positive	77
True Positive	9,991
Total Correct Predictions	19,914
% of Correct Prediction	99.57%
Total Incorrect Predictions	86
% of Incorrect Predictions	0.43%

This model also excelled with almost perfect accuracy, although errors seem to be slightly skewed towards 77 false positives versus 9 false negatives, suggesting a degree of sensitivity towards the high-revenue class. Nevertheless, the model still performed optimally and still reliable but compared to XGBoost it shows less balanced error distribution so it is marginally less stable.

Figure 38.

(a) Confusion Matrix for Enhanced Neural Network



(b) Confusion Matrix Interpretation

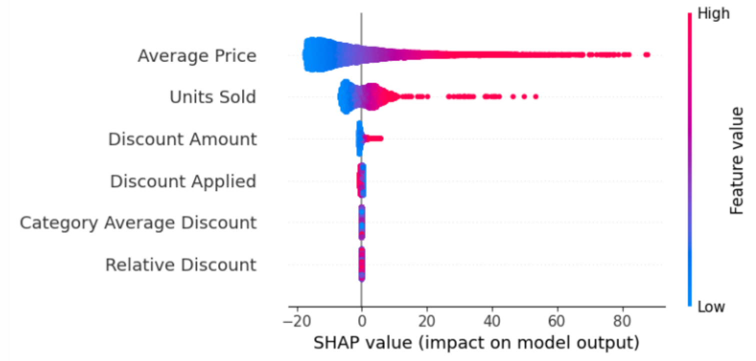
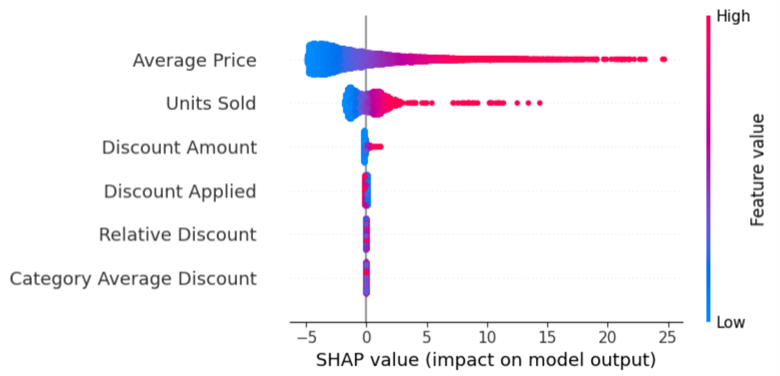
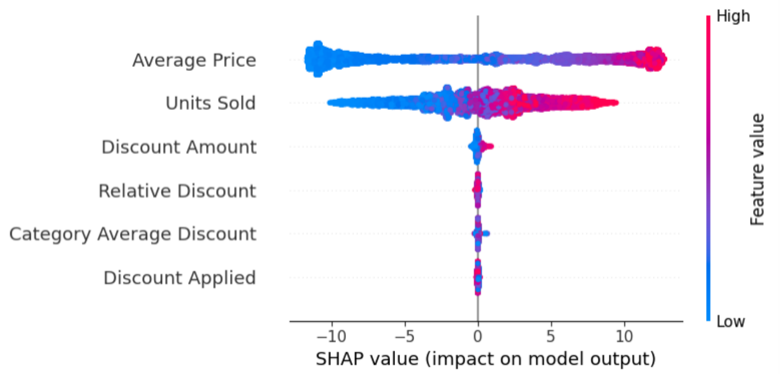
Number of Instances	20,000
True Negative	9,974
False Negative	69
False Positive	26
True Positive	9,931
Total Correct Predictions	19,905
% of Correct Prediction	99.53%
Total Incorrect Predictions	95
% of Incorrect Predictions	0.48%

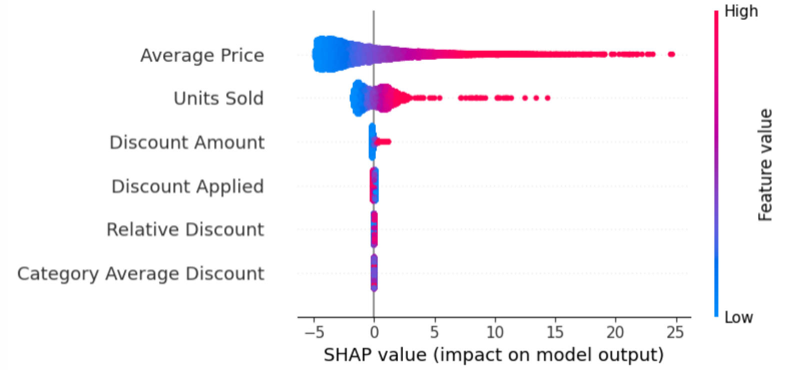
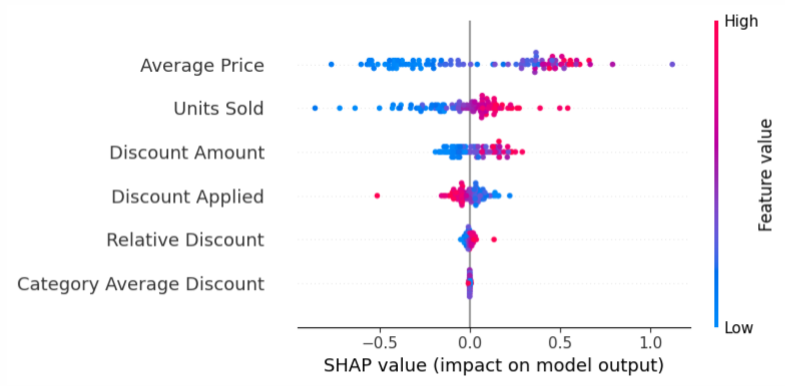
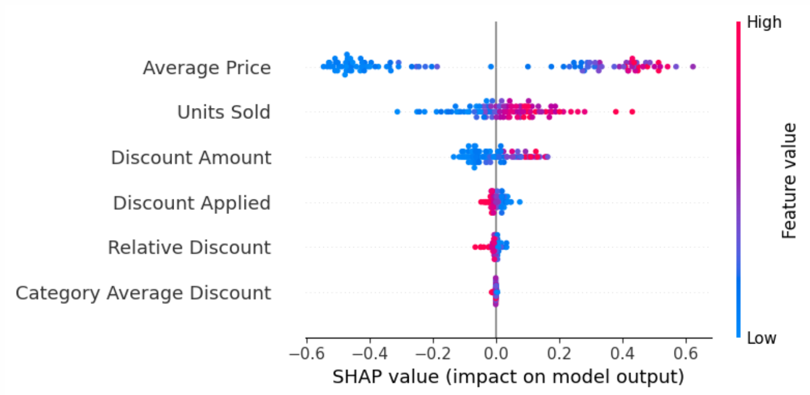
The Enhanced Neural Network performed similarly to the Simple Neural Network. It demonstrated great precision for high-revenue predictions given only 26 false positives, but resulted in slightly lower recall compared to XGBoost, with 69 false negatives. Despite this, overall performance was still exceptional.

7.3.3 *Further Evaluation Metrics + SHAP for Framing Bias*

Table 25.

Table comparing training time, generalisation (overfit gap) and most influential features in social proof bias-induced purchases across models

Model	Training Time (s)	Overfit Gap	Top Feature
Logistic Regression	1.86	0.002	
Linear SVM	1.61	0.002	
XGBoost	20.20	0.002	

Random Forest	68.69	0.004	
Simple Neural Network	217.28	0.000	
Enhanced Neural Network	482.10	0.000	

SHAP analysis revealed “Average Price” and “Units Sold” were the most dominant features across all models, while variables directly feature engineered to reflecting framing effects like “Discount Amount” and “Relative Discount” had less influence. Thus, in this dataset, price sensitivity and sales volume had more influence on customer purchasing behaviour than discount framing signals; albeit the feature engineered variable did contribute, but their weaker significance suggests the models focused on other motivators besides psychological framing effects.

Further performance metrics revealed Logistic Regression and Linear SVM train very quickly with negligible overfit gaps, but they offer less sophisticated feature weighting.

7.4 Cross Bias performance comparison

Table 26.
Model performance metrics comparison across all models and all psychological biases

	Precision (macro average)	**Recall** (macro average)	**F1-Score** (macro average)	**Number of Instances**	**Total Correct Predictions**	**Total Incorrect Predictions**	**Accuracy**	**Overfit Gap**	**Training Time (s)**
Urgency Bias									
Logistic Regression	0.95	0.79	0.85	37,015	34,054	2,961	92.00%	0.003	1.40
Linear SVM	0.95	0.79	0.85	37,015	34,054	2,961	92.00%	0.003	0.26
XGBoost	0.94	0.80	0.85	37,,015	34,033	2,982	91.94%	0.000	0.89
Random Forest	0.86	0.82	0.90	37,015	32,328	4,687	95.10%	0.049	12.09
Simple Neural Network	0.95	0.79	0.84	55,522	50,530	4,992	91.67%	0.003	407.03
Enhanced Neural Network	0.95	0.79	0.84	55,522	48,041	4,481	91.67%	0.003	643.73
Social Proof Bias									
Logistic Regression	0.76	0.53	0.51	4,698	3.889	809	82.64%	0.001	0.30
Linear SVM	0.80	0.52	0.50	4,698	3.891	807	82.68%	0.002	0.18
XGBoost	0.67	0.52	0.49	4,698	3,864	824	82.81%	0.006	0.46
Random Forest	0.67	0.52	0.49	4,698	3,865	833	82.84%	0.006	3.02
Simple Neural Network	0.80	0.52	0.50	4,698	3,890	808	82.68%	0.002	81.38
Enhanced Neural Network	0.81	0.52	0.50	4,698	3,892	806	82.67%	0.002	65.21
Framing Bias									
Logistic Regression	0.95	0.95	0.95	19,694	18,754	940	95.09%	0.002	1.86
Linear SVM	0.95	0.95	0.95	19,700	18, 752	948	95.06%	0.002	1.61
XGBoost	1.00	1.00	1.00	20,000	19,950	50	99.94%	0.002	20.20
Random Forest	1.00	1.00	1.00	20,000	19,927	73	100.00%	0.004	68.69
Simple Neural Network	1.00	1.00	1.00	20,000	19,914	86	99.65%	0.001	217.18
Enhanced Neural Network	1.00	1.00	1.00	20,000	19,905	95	99.54%	0.000	482.10

When comparing across the three biases, framing bias appears the most predictable with every model exceeding 95% accuracy. Ensemble techniques performed the strongest achieving 100% accuracy for the Random Forest model. Nevertheless, linear models still performed well, as did neural networks which matched the performance of ensemble methods, but at greater computational power, making them less efficient despite having strong predictive power.

Urgency bias also performed strongly with models consistently performing at approximately 92% with F1-Scores around 0.85. There is no clear model dominating – Logistic Regression, SVM and XGBoost performed similarly and consistently, balancing both recall and precision scores well, unlike Random Forest which may have had the highest accuracy but was less balanced in performance with the best recall (0.82) but at the expense of precision (0.86). Neural networks took significantly longer to train without meaningfully improving the results, deeming it the least practical option for this bias.

Social proof bias was the most difficult to predict, accuracies plateaued to 82-83% and F1-scores near 0.50 across all models. There was no clear indication of any superior models. Linear models and the neural networks resulted in relatively higher precision scores (~0.80), while ensemble methods performed only marginally weaker. Precision scores were relatively higher than recall suggesting models performed better in detecting strong social proof signals than weaker or noisier patterns in the data. The review-based dataset used for this bias is based on consumer opinion data, introducing a lot of variability and noise to the dataset.

Across all three biases, ensembles including XGBoost and Random Forest outshone other models for framing bias and linear models including Logistic Regression and SVM for urgency; nevertheless, these models only slightly outperformed other models so they are only marginally superior. No model dominated for social proof. Neural networks performed consistently well, but across all models they are considered the least practical. They took significantly longer to train, without any noticeable gains compared to even baseline models.

Overall, there is no clear model that was superior, but Random Forest offered the best balance across all metrics for all biases. These findings imply model effectiveness is very bias-dependent, where models may perform better in some biases or weaker in others – structured datasets (framing bias) aligned with ensemble methods, data focusing on temporal patterns (urgency bias) favoured linear efficiency, while noisy opinion-based data (social proof) limited model differentiation.

7.5 Chapter Summary

This chapter compared model performance within and across biases through precision, recall, F1-score, accuracy, overfitting gap and training time metrics. There were marginal differences between algorithms, although, Random Forest was the most consistent in offering the most balance performance overall, while neural networks provided little value given longer training times with no performance impact. Linear models acted as strong, interpretable baselines, performing similarly to ensemble methods. Engineered features were not always influential, particularly engineered framing bias features, in predictive outcomes, highlighting the impact of data quality rather than model complexity. Nevertheless, urgency features and some social proof bias features influenced the predictions, aligning with psychological theory.

Chapter 8:

Discussion + Evaluation Against Project Requirements

8.1 Interpretation of results + Critical Evaluation

This project aimed to determine if consumer behaviour could be predicted where urgency, social proof and framing bias features influence purchase likelihood. Using SHAP, it examines whether these psychological constructs meaningfully impact the predictions. Findings revealed there is a clear hierarchy in detecting biases: purchases influenced by discount framing were the easiest to predict, those by urgency were moderately predictable while purchases driven by social proof were the hardest to predict. Performance varied marginally across models, but these changes were less significant than the type and quality of the datasets, showing how computational bias detection is more dependent on representation and ecological validity than complex algorithms.

Framing bias resulted in near flawless predictions across all models. This may imply framing effects are simpler to capture, aligning with González et al. (2016) work on how discount framing influences decision-making. Although, the dataset used for this bias was synthetic and very structured, resulting in clean and balanced data. This raises an ecological validity concern since framing effects in the real world interact with other factors such as competing prices against competitors or context for discounting products which makes them less predictable. Perhaps purchases were easier to predict with discount-framing influence because of dataset design rather than the intrinsic quality of framing.

Urgency bias related purchases were moderately easy to predict, not as much as discount-framed purchases. With high accuracy but low recall, purchases identified as urgent were often misclassified as non-urgent – a pattern consistent with class imbalance. Class imbalance was acknowledged during data preparation but not rectified since it reflects real-world purchases (urgency driven purchases are not an everyday or regular occurrence). Urgent instances became the minority, skewing predictions towards the majority, this is a well-known limitation recognised in machine learning literature (He & Garcia, 2009), implying that models cannot completely overcome structural imbalance. Random Forest results emphasise the trade-off between sensitivity and precision as recall increased but also produced more false positives. Therefore, this reflects the situational nature of urgency bias from a psychological perspective; while it is powerful in some settings (Cialdini, 2001), urgency signals are inconsistent and hard to integrate into features for predictive modelling – depleting model dependability.

Social Proof induced purchases were the weakest, with findings remaining consistent regardless of the model choice. No level of model complexity was able to compensate for the noisy nature of review-based data. Although precision outperformed recall, models failed to recognise where social proof was weak or missing, aligning with precious results that reviews are prone to herding effects, positivity skew and self-selection bias (Hu et al., 2009). Hence, errors in this instance seem due to underlying data rather than methodological issues.

All in all, two major findings emerge. Firstly, dataset structure was critical: near-perfect results were achieved where data is thoroughly cleaned and synthetic (framing bias), but where there is class imbalance, performance is hindered (urgency bias) and unpredictable factors such as social proof can affect accuracy. Secondly, model complexity had little effect. Neural networks did not significantly outperform simple models despite its longer training times, supporting Domingos' (2012) claim that strong data is more important than sophisticated, complex methods.

8.1 Model-Level Insights

Across all three biases, no single model consistently dominated, rather, performance reflected the interplay between dataset design and model type. Linear models (Logistic Regression and Linear SVM), ensemble models (Random Forest and XGBoost), neural networks (Simple and Enhanced Neural Network) all had advantages and disadvantages, but the overall results confirm algorithm complexity did not better results.

Linear models proved very efficient, excelling in both framing and urgency bias tasks with little training times and great generalisation. They serve as valuable baselines with their consistency and interpretability; however, they were less capable of detecting minor non-linear interactions which explains their struggle in noisier contexts such as detecting purchases influenced by social proof.

Ensemble models had the strongest predictive ability overall, obtaining near-perfect accuracy, recall and F1-score, where Random Forest also enhanced urgent recall, catching more minority class situations. Although, Random Forest showed signs of overfitting in detecting urgency-induced purchases, but this is only relative to other overfitting gap values, it is still a negligible value. Neither ensemble technique was able to overcome the challenges of the review-based social proof features, implying ensembles perform well in structured environments but do not address basic issues of data variability.

Neural networks closely resembled the performance of ensemble models but at greater computational cost. Training time were drastically longer with negligible performance gains. These networks can capture deeper, non-linear interactions but in this study, they did not outperform ensembles or linear models – indicating that data quality was the most important factor in predictive success.

Table 27 summarises the strengths and weaknesses of the three model families.

Table 27.
Model-Level Insights Across Biases

Model Family	Strengths	Weaknesses	Role in the Project
Linear Models	Fast Training Interpretable	Struggled with non-linear interactions and limited performance in noisy contexts like social proof.	Provides an interpretable baseline, showing how simple models can outperform complex ones when data is well-structured.
Ensemble Models	Strongest predictive performance overall, near perfect in framing bias.	Most likely to overfit and could not overcome noisy review data for social proof induced purchases.	Best balance in accuracy and generalisation, supports the value of ensemble methods in structured datasets.
Neural Networks	Can handle complex, non-linear interactions	Very computationally expensive; significantly increased training times without any distinct performance gains over ensemble methods.	Showed the little benefit of added complexity, reinforcing the importance of data quality than model complexity.

8.2 Feature Engineering Value + SHAP Analysis

Traditional metric values (precision, recall, F1-score and accuracy) reflect the predictive power of models but SHAP analysis was mandatory in assessing if the engineered features operationalising urgency, social proof and framing bias influence predictions in manners aligning with psychological theory. Having an interpretability step like so is critical to the project's aim of addressing the gap between machine learning and behavioural science, revealing if models could learn meaningful bias-related trends.
Social proof also showed the strongest influence in predicting social proof induced purchases, but the engineered features had meaningful influence on model predictions. SHAP ranked "Average Rating Per Item," "Total Helpfulness Per Item", and the most expected influential "Social Proof Score" as the three top

contributors across all models. Higher ratings and social influence increased the likelihood of purchase recommendations, highlighting models are sensitive to peer influence cues. This aligns with Rosario et al. (2016) findings that societal positive feedback influences purchase decisions, capturing social proof bias at play. Therefore, unlike framing features, social proof proxies were not overshadowed by raw transactional variables but actually played a central role in recommended classification. Nevertheless, the social proof dataset performed modestly with approximately 82-83% accuracy, so even though features had influence, they somewhat struggled in classifying both recommended and not-recommended instances – reflecting the noise of review data where online reviews are subjective and prone to self-selection and skewness due to herding effects (Hu et al. 2009), limiting its consistency as predictive cues. Overall, this confirms via SHAP that social proof features have great explanatory power, but their predictive accuracy was limited by data variability instead of feature irrelevance.

Urgency features also demonstrated strongest influence of urgency cues on purchases. SHAP confirmed that temporal spikes such as "Days to Nearest Holidays" increased the probability of urgent classification, consistent with scarcity research showing urgency effect is heightened when deadlines loom and prompt quicker decision-making (Cialdini, 2001). Although weaker urgency cues such as "Holiday Purchase" had little influence in urgent classification, but this is understandable since most urgent purchases are not made on the day of holidays itself but in days leading up to it where time pressure is most prominent (Sun et al., 2023), hence "Days to Nearest Holiday" had greater impact. Therefore, SHAP proved all models aligned with psychological expectations where time-sensitive cues influenced purchase decisions.

Framing features was not as strong as urgency features in predicting high revenue purchases. Variables such as "Discount Amount," "Discount Applied," Category Average Discount," and "Relative Discount" were feature engineered intending to capture framing effects but their SHAP feature importance revealed only a minor influence. Predictors such as "Average Price" (engineered for purpose of other features not to reflect high value purchases) and "Units Sold" (variable from the raw dataset, not engineered) dominated feature influence. Thus, while framing bias is powerful in theory, proxies used here failed to replicate its impact in practice; this highlights a limitation of feature engineering where raw dataset variable can overshadow more subtle engineered cues.

In summary, SHAP analysis revealed engineered bias features were not consistently successful. Social proof features had great impact, as did some urgency features, but they were often overshadowed by other transactional variables, particularly for framing features. This portrays the potential and limitations of bias-aware feature engineering. Psychological constructs can be embedded into predictive modelling, but its efficacy depends on how strong the proxies are and if the datasets are well-organised and structured.

8.3 Theoretical & Practical Implications

Theoretically, this project advances the field of demonstrating psychological biases can be operationalised as engineered features and tested using SHAP to determine if it can align with behavioural theory. By going beyond generic consumer prediction, this study adds new insight towards bias-aware modelling where social proof proxies provided strong signals while urgency and framing were somewhat restrained by class imbalance and data noise, respectively. Therefore, this project's findings highlight that computational models can accurately reflect psychological constructs, provided the proxies and datasets are reliable.

These results have value for multiple stakeholders in practical settings. Policymakers can employ bias-aware algorithms to audit digital platforms and monitor manipulative persuasive design. Businesses can use this concept responsibly to improve transparency, showing how reviews, discounts and time-sensitive deadlines actually affect sales. This allows businesses, particularly marketing companies to maximise marketing strategies for product sales, responsibly and ethically.

8.4 Evaluation Against Project Requirements

The four core requirements of this project were successfully addressed. *Bias-aware feature engineering* was accomplished by encoding urgency, social proof and framing bias as features. However, their impact varied: social proof proxies were strong where engineered features consistently drove predictions, even though overall accuracy was hindered by noisy data. Conversely, framing proxies had the highest performance across metrics but models relied on raw transaction variables such as "Units Sold' rather than the engineered features, raising concern if discount-based features are adequate and applicable. Urgency proxies were able to highlight temporal spikes influencing purchase decisions due to time-sensitivity, but it struggled with more subtle cues perhaps due to the class imbalance in the dataset. *Multi-model evaluation* revealed dataset quality is more important than model type since performance difference between linear, ensemble and neural network models were very minimal. *Interpretability* was achieved through using SHAP and validated these results, emphasising social proof features influence in recommendations, and urgency features in time-sensitive purchases, but also framing features' inability to drive purchase decisions.

8.5 Chapter Summary

This chapter critically analysed results across all biases. Results conveyed that dataset structure influenced results more than model choices. SHAP analysis was integral to this project, in confirming if engineered features aligned with psychological theory or not to bridge the gap between behavioural science and predictive modelling. Future work was suggested to better feature design, model interactions and dataset preparation.

Chapter 9:
Conclusion

9.1 Summary

This dissertation investigated where psychological biases may be built into machine learning models of consumer purchase behaviour using feature engineering and interpretability tools (SHAP). Across three datasets, urgency, social proof and framing biases were operationalised as specific features and tested using linear, ensemble and neural network models. This was followed by SHAP analysis to determine if these features affected predictions in a manner that is compatible with behavioural theory.

The investigation discovered that engineered proxies did not perform evenly. Urgency related purchases were influenced by temporal effects, but more subtle cues struggled to have any impact. Social proof induced purchases were consistently influenced by engineered features such as ratings and social proof scores although it's performance metrics were relatively moderate with the lowest accuracies across all biases. Despite the strongest performance metrics, framing relevant features contributed little to high revenue purchase predictions, where raw variables were more influential.

The findings overall implied data quality and feature design was more integral to bridging the gap between behavioural theory and predictive modelling than model choice, emphasising the need for careful, theory-driven feature engineering.

9.2 Addressing Research Questions

This study was led by four research questions, all of which can now be addressed:

RQ1: ***Can machine learning detect psychological biases when engineered into datasets?***
Yes, but unevenly. Urgency and social proof characteristics meaningfully impacted predictions, while high revenue purchase predictions were overshadowed by raw dataset variables for framing bias.

RQ2: ***Which models perform most effectively across biases?***
There was no single superior model family. XGBoost and Random Forest were the strongest performers in structured framing data but in comparison to other models, predictive performance was only marginally better. Linear models (logistic regression and linear SVM) proved efficiency in urgency bias tasks and strong interpretable baselines across framing and social proof bias tasks, while neural networks provided no performance enhancement with their computational cost across all biases.

RQ3: ***What features most strongly drive model predictions, and do they align with psychological expectations?***
SHAP analysis revealed temporal factors (urgency feature) and ratings/social proof scores (social proof features) aligned with behavioural theory. However, framing features had weaker contributions and highlighted the criticality of feature design.

9.3 Future Work

Firstly, future work should improve the feature engineering conducted, particularly for framing bias. The engineered discount-relevant proxies had little impact compared to raw variables such as average price and units sold. Another approach could use relative price framing across competing products, for example, to better capture a cognitive bias influence.

Secondly, in real purchasing contexts, biases are likely to interact, yet the datasets used here treated them separately; for instance, urgency may enhance the persuasive power of social proof, or discount framing may amplify the effect of urgency. Although this project examined each bias alone to test their individual predictive value, future research should consider multi-bias feature engineering to better the ecologically valid insights into real-world consumer decision-making.

Finally, the quality of data is critical. Social proof engineered features proved to be strong, but its noisy review data restricted its predictive performance. A possible solution for future work could utilise natural language processing of review texts instead of simply dropping this feature, alongside ratings and helpfulness scores to improve signal strength.

9.4 Critical reflection

The novelty of this dissertation involved incorporating psychological theory into machine learning pipelines using bias-aware feature engineering and SHAP interpretation. A crucial strength was showcasing how engineered features for urgency and social proof can influence predictions in ways consistent with theory, while simultaneously highlighting areas where proxies for framing fell short. This revealed new insight into the challenge of uncovering subtle behavioural processes when powerful transactional features have greater influence.

This study did have some limitations. Ecological validity across biases was uneven where synthetic data used for framing bias simplified real-world complexity. Although, the social proof dataset was skewed and noisy, reducing its dependability. It was expected that performance metrics and interpretability would be enhanced depending on model choice, but feature design appeared as more decisive and remains a critical methodological problem.

Overall, the project provides a benchmark for linking behavioural science and predictive modelling, however its findings emphasise the importance of careful feature design and better data source and selection for bias-aware analytics to achieve both academic and practical value.

Chapter 10:
References

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, *58*, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Askalidis, G., & Malthouse, E. C. (2016). Understanding and overcoming biases in customer reviews. *arXiv preprint arXiv:1604.00417*. https://arxiv.org/abs/1604.00417

Azad, M. S., Khan, S. S., Hossain, R., Rahman, R., & Momen, S. (2023). Predictive modeling of consumer purchase behavior on social media: Integrating theory of planned behavior and machine learning for actionable insights. *Plos one*, *18*(12), e0296336. https://doi.org/10.1371/journal.pone.0296336

Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of marketing research*, *53*(3), 297-318. https://doi.org/10.1126/science.1240466

Balyemah, A. J., Weamie, S. J., Bin, J., Jarnda, K. V., & Joshua, F. J. (2024). Predicting Purchasing Behavior on E-Commerce Platforms: A Regression Model Approach for Understanding User Features that Lead to Purchasing. *International Journal of Communications, Network and System Sciences*, *17*(6), 81-103. https://doi.org/10.4236/ijcns.2024.176006

Bastos, J. A., & Bernardes, M. I. (2024). Understanding online purchases with explainable machine learning. *Information*, *15*(10), 587. https://doi.org/10.3390/info15100587

Baydaş, A., Ata, S., & KÖK, N. (2020). An Empirical Study to Determine the Impact of Black Friday Days on Consumer Purchasing Behavior, *1*(2), 87.

Boyd Thomas, J., & Peters, C. (2011). An exploratory investigation of Black Friday consumption rituals. *International journal of retail & distribution management*, *39*(7), 522-537. https://doi.org/10.1108/09590551111144905

Chen, Y., Liu, H., Wen, Z., & Lin, W. (2023). How explainable machine learning enhances intelligence in explaining consumer purchase behavior: A random forest model with anchoring effects. *Systems*, *11*(6), 312. **https://doi.org/10.3390/systems11060312**

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, *43*(3), 345-354. https://doi.org/10.1509/jmkr.43.3.345

Cialdini, R. B. (2001). The science of persuasion. *Scientific American*, *284*(2), 76-81. https://doi.org/10.1038/scientificamerican0201-76

Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika*, *72*(2), 263-268. https://doi.org/10.1007/s11336-006-1560-6

Cui, A. (2022). *E-commerce behavior data from multi-category online store* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/YGLYDY

Dinata, I. M., Sinaga, V. L., & Wibowo, A. (2025). Comparing multiple ensemble classifiers for e-commerce recommendation system. *International Journal of Engineering Trends and Technology, 73*(1), 250–254. https://doi.org/10.14445/22315381/IJETT-V73I1P121

Domingos, P. (2012). *A few useful things to know about machine learning. Communications of the ACM, 55*(10), 78–87. https://doi.org/10.1145/2347736.2347755

Esmeli, R., & Gokce, A. (2025). An Analysis of Consumer Purchase Behavior Following Cart Addition in E-Commerce Utilizing Explainable Artificial Intelligence. *Journal of Theoretical and Applied Electronic Commerce Research*, *20*(1), 28. https://doi.org/10.3390/jtaer20010028

Godinho, S., Prada, M., & Garrido, M. V. (2016). Under pressure: An integrative perspective of time pressure impact on consumer decision-making. *Journal of International Consumer Marketing*, *28*(4), 251-273. https://doi.org/10.1080/08961530.2016.1148654

González, E. M., Esteva, E., Roggeveen, A. L., & Grewal, D. (2016). Amount off versus percentage off—when does it matter?. *Journal of Business Research*, *69*(3), 1022-1027. https://doi.org/10.1016/j.jbusres.2015.08.014

Grewal, D., Krishnan, R., Baker, J., & Borin, N. (1998). The effect of store name, brand name and price discounts on consumers' evaluations and purchase intentions. *Journal of retailing*, *74*(3), 331-352. https://doi.org/10.1016/S0022-4359(99)80099-2

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, *21*(9), 1263-1284. https://doi.org/10.1109/TKDE.2008.239

Hu, N., Zhang, J., & Pavlou, P. A. (2009). *Overcoming the J-shaped distribution of product reviews*. *Communications of the ACM, 52*(10), 144–147. https://doi.org/10.1145/1562764.1562766

Kahneman, T. (1979). D. kahneman, a. tversky. *Prospect theory: An analysis of decisions under risk*, 263-291. https://doi.org/10.2307/1914185

Lang, T., & Rettenmeier, M. (2017, April). Understanding consumer behavior with recurrent neural networks. In *Workshop on Machine Learning Methods for Recommender Systems*.

Lin, J. (2025). Application of machine learning in predicting consumer behavior and precision marketing. *PLoS One*, *20*(5), e0321854. https://doi.org/10.1371/journal.pone.0321854

Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, *341*(6146), 647-651. https://doi.org/10.1126/science.1240466

Muzumdar, P. (2021). Impact of review valence and perceived uncertainty on purchase of time-constrained and discounted search goods. *arXiv preprint arXiv:2110.09313*. https://doi.org/10.48550/arXiv.2110.09313

Nicolaou, N. (2018). *Women's e-commerce clothing reviews* [Data set]. Kaggle. https://www.kaggle.com/datasets/nicapotato/womens-ecommerce-clothing-reviews

Noti, G., Levi, E., Kolumbus, Y., & Daniely, A. (2016). Behavior-based machine-learning: A hybrid approach for predicting human decision making. *arXiv preprint arXiv:1611.10228*. https://doi.org/10.48550/arXiv.1611.10228

Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and translational science*, *17*(11), e70056. https://doi.org/10.1111/cts.70056

Raiqal, M. G., & Mukaram, M. (2025). The influence of seasonal price discounts on online impulsive buying decisions on the Steam gaming platform (Study on Generation Z in Bandung), *6*(5), 3104–3111. https://doi.org/10.38035/jemsi.v6i5.5050

Requena, B., Cassani, G., Tagliabue, J., Greco, C., & Lacasa, L. (2020). Shopper intent prediction from clickstream e-commerce data with minimal browsing information. *Scientific reports*, *10*(1), 16983. https://doi.org/10.1038/s41598-020-73622-y

ReviewTrackers. (2023, April 27). *What is social proof? Definition, examples, and types.* ReviewTrackers. https://www.reviewtrackers.com/blog/social-proof/

Shah, I. A. (2023). *Comprehensive synthetic e-commerce dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/imranalishahh/comprehensive-synthetic-e-commerce-dataset

Shi, X., Li, F., & Chumnumpan, P. (2020). The use of product scarcity in marketing. *European Journal of Marketing*, *54*(2), 380-418. https://doi.org/10.1108/EJM-04-2018-0285

Sun, Y., Li, Y., & Zhou, J. (2023). Relationship between time pressure and consumers' impulsive buying: Role of perceived value and emotions. *Frontiers in Psychology, 14,* 1250205. https://doi.org/10.3389/fpsyg.2023.1250205

Suri, R., & Monroe, K. B. (2003). The effects of time constraints on consumers' judgments of prices and products. *Journal of consumer research*, *30*(1), 92-104. https://doi.org/10.1086/374696

Tagliabue, J., Lacasa, L., Greco, C., Pavoni, M., & Polonioli, A. (2019). Predicting e-commerce customer conversion from minimal temporal patterns on symbolized clickstream trajectories. *arXiv preprint arXiv:1907.02797*. https://arxiv.org/abs/1907.02797

Thaler, R. H. (1985). *Mental accounting and consumer choice*. *Marketing Science, 4*(3), 199–214. https://doi.org/10.1287/mksc.4.3.199

Thomas, M., & Morwitz, V. (2005). How Do Consumers and Managers Process Numeric Information? The Role of Numerical Cognition. *Advances in Consumer Research*, *32*, 445.

Tokuç, A. A., & Dağ, T. (2025). Predicting User Purchases From Clickstream Data: A Comparative Analysis of Clickstream Data Representations and Machine Learning Models. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3548267

Aggarwal, P., & Vaidyanathan, R. (2003). Use it or lose it: purchase acceleration effects of time-limited promotions. *Journal of Consumer Behaviour: An International Research Review*, *2*(4), 393-403. https://doi.org/10.1002/cb.116

Vallarino, D. (2023). Buy when? Survival machine learning model comparison for purchase timing. *arXiv preprint arXiv:2308.14343*. https://doi.org/10.48550/arXiv.2308.14343

Van den Poel, D., & Buckinx, W. (2005). Predicting online-purchasing behaviour. *European journal of operational research*, *166*(2), 557-575. https://doi.org/10.1016/j.ejor.2004.04.022

Verma, R., Rathor, D., Kumar, S., Mishra, M., & Baranwal, M. (2025). Enhancing Customer Repurchase Prediction: Integrating Classification Algorithms with RFM Analysis for Precision and Actionable Insights. *IIMB Management Review*, 100574. https://doi.org/10.1016/j.iimb.2025.100574

Wang, W., Xiong, W., Wang, J., Tao, L., Li, S., Yi, Y., ... & Li, C. (2023). A user purchase behavior prediction method based on xgboost. *Electronics*, *12*(9), 2047. **https://doi.org/10.3390/electronics12092047**

Yang, C. Y., & Chen, Y. Z. (2024). Support vector machine classification of patients with depression based on resting-state electroencephalography. *Asian Biomedicine: Research, Reviews and News*, *18*(5), 212-223.

Yixuan, Z. (2024). Utilizing machine learning algorithms for consumer behaviour analysis. *Applied and Computational Engineering*, *49*, 213-219. https://doi.org/10.54254/2755-2721/49/20241186

Zhang, G., Li, J., Tan, M., & Zhong, Y. (2023). The influence of green product type, message framing, and anticipated pride on green consumption behavior: An event-related potential (ERP) study. *Brain Sciences*, *13*(10), 1427. https://doi.org/10.3390/brainsci13101427

Chapter 11:
Appendices

11.1 Signed Project specification form

Figure 39.

Project Specification Form

Project Specification (Insert your full name)

Project Specification

A. Basic details

Student name:	Davina Khimani
Draft project title:	Can Machine Learning Detect Psychological Biases in Consumer Purchasing Behaviour? A Behavioural Data Approach to Predictive Modelling and Interpretation
Course and year:	MSc Data Analytics (2025-2026)
Client organisation:	
Client contact name:	
Project supervisor:	Dr Ella Haig

B. Outline of the project environment

This project combines psychology/marketing (consumer behaviour) and machine learning. It will use an e-commerce or retail dataset to predict purchase behaviour based on psychological biases that will be feature engineered to represent loss aversion, anchoring effect and social proof.

C. The problem to be solved

Traditional ML models can predict consumer purchase behaviour but fail to account for underlying psychological biases. This project attempts to enhance predictive performance by incorporating features based on psychological biases, providing a clearer understanding of how such biases influence consumer behaviour.

D. Breakdown of tasks

D.1. Approach

- Conduct background research on consumer psychology (loss aversion, social proof, anchoring)
- Identify a suitable dataset
- Engineer features from behavioural data to represent psychological biases
- Build and evaluate machine learning models using skills learnt from modules
- Analyse model interpretability using SHAP/feature importance

D.2. Background Research

- Behavioural economics literature
- ML techniques for behavioural predictions
- Feature engineering strategies for psychological concepts

D.3. Tools required

- Python

Printed: 22 May 2025 Page: 2

- Google Collab

D.4. Skills needed

- Machine Learning Modelling
- Psychology knowledge

E. Project deliverables

- Trained machine learning models predicting consumer purchases
- Engineered dataset with psychological bias indicators
- Model analysis report – feature importance/SHAP
- Dissertation document

F. Requirements

N/A

G. Legal, ethical, professional, social issues

Data Privacy: only using anonymous public datasets

No other ethical issues

H. Facilities and resources

Personal Laptop/PC for Google Collab containing python and machine learning libraries. Datasets from public websites. No constraints anticipated.

I. Project plan

Commencing from 2nd June 2025:
02/06 – 16/06: Background research on biases and dataset search

16/06-23/06: Dataset selection and preprocessing

23/06 – 30/06: Feature engineering to model psychological biases and ML model building

30/06-07/07: Write up – Introduction

07/07-14/07: Write up – Literature review

14/07-21/07: Write up – data preprocessing/feature engineering and model building

21/07-28/07: Write up – results

28/07-04/08: Write up – discussion + conclusion

J. Supervision meetings

After each major milestone there will be a meeting, for example, once a dataset has been decided another meeting will be scheduled to confirm the dataset and methodology. Then the following meetings will be scheduled for when the code is completed before writing up to adjust any issues or errors. Followed by meetings after sections of the report have been reviewed by supervisor.

K. Project mode

If there are two possibilities for your project mode, after negotiation, please record your planned duration and submission date. It is also helpful to record your initial registration mode (i.e. are you a full time or a part time student). Remember, the exact dates will be announced through Moodle – these represent a generic guideline.

	Please delete as appropriate	
Registration mode	Full Time	
Project mode	Full Time	
Planned submission deadline	11/9/20	

L. Signatures

	Signature:	Date:
Student		22/05/2025
Client		
Project supervisor		

11.2 Ethics approval certificate

Figure 40.

Ethics Review Certification

UNIVERSITY OF PORTSMOUTH

Certificate of Ethics Review

Project title: Modelling the Impact of Psychological Biases on Consumer Choice Using Machine Learning

Name:	Davina Khimani	**User ID:**	2089114	**Application date:**	02/06/2025 16:51:59	**ER Number:**	TETHIC-2025-111199

You must download your referral certificate, print a copy and keep it as a record of this review.

The FEC representative(s) for the **School of Computing** is/are **Elisavet Andrikopoulou, Kirsten Smith**

It is your responsibility to follow the University Code of Practice on Ethical Standards and any Department/School or professional guidelines in the conduct of your study including relevant guidelines regarding health and safety of researchers including the following:

- University Policy
- Safety on Geological Fieldwork

It is also your responsibility to follow University guidance on Data Protection Policy:

- General guidance for all data protection issues
- University Data Protection Policy

Which school/department do you belong to?: **School of Computing**
What is your primary role at the University?: **Postgraduate Student**
What is the name of the member of staff who is responsible for supervising your project?: **Dr Ella Haig**
Is the study likely to involve human subjects (observation) or participants?: No
Will financial inducements (other than reasonable expenses and compensation for time) be offered to participants?: No
Are there risks of significant damage to physical and/or ecological environmental features?: No
Are there risks of significant damage to features of historical or cultural heritage (e.g. impacts of study techniques, taking of samples)?: No
Does the project involve animals in any way?: No
Could the research outputs potentially be harmful to third parties?: No
Could your research/artefact be adapted and be misused?: No
Will your project or project deliverables be relevant to defence, the military, police or other security organisations and/or in addition, could it be used by others to threaten UK security?: No
Please read and confirm that you agree with the following statements: I confirm that I have considered the implications for data collection and use, taking into consideration legal requirements (UK GDPR, Data Protection Act 2018 etc.), I confirm that I have considered the impact of this work and and taken any reasonable action to mitigate potential misuse of the project outputs, I confirm that I will act ethically and honestly throughout this project

Supervisor Review

As supervisor, I will ensure that this work will be conducted in an ethical manner in line with the University Ethics Policy.

Supervisor comments:

Supervisor's Digital Signature: **ella.haig@port.ac.uk** Date: **04/07/2025**

11.3 Code for Model Building

11.3.1 Logistic Regression

Figure 41.

Code for Logistic Regression

```
import pandas as pd
import numpy as np
import time
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, confusion_matrix
import shap

X = df.drop(columns=["Category", "High Revenue", "Revenue"])
y = df["High Revenue"]

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, stratify=y, random_state=42)

start_time = time.time()
model = LogisticRegression(max_iter=1000)
model.fit(X_train, y_train)
train_time = time.time() - start_time

y_train_pred = model.predict(X_train)
y_test_pred = model.predict(X_test)

# Classification report
report_df = pd.DataFrame(
    classification_report(y_test, y_test_pred, output_dict=True))

# Overfitting gap
train_acc = model.score(X_train, y_train)
test_acc = model.score(X_test, y_test)
overfit_gap = abs(train_acc - test_acc)

# Confusion Matrix
cm = confusion_matrix(y_test, y_test_pred)
plt.figure(figsize=(5, 4))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.title("Confusion Matrix - Logistic Regression (Framing Bias)")
plt.show()

# SHAP
explainer = shap.LinearExplainer(model, X_train, feature_perturbation="interventional")
shap_values = explainer.shap_values(X_test)
shap.summary_plot(shap_values, X_test, feature_names=X.columns)

print("\nLogistic Regression Evaluation - Framing Bias Dataset")
print(f"Training Time: {train_time:.4f} seconds")
print(f"Train Accuracy: {train_acc:.4f}")
print(f"Test Accuracy: {test_acc:.4f}")
print(f"Overfitting Gap: {overfit_gap:.4f}")
display(report_df)
```

11.3.2 Support Vector Machine

Figure 42.
Code for Support Vector Machine

```
import pandas as pd
import numpy as np
import time
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.svm import LinearSVC
from sklearn.metrics import classification_report, confusion_matrix
import shap


X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, stratify=y, random_state=42)

start_time = time.time()
model = LinearSVC(max_iter=5000, random_state=42)
model.fit(X_train, y_train)
train_time = time.time() - start_time

y_train_pred = model.predict(X_train)
y_test_pred = model.predict(X_test)

# Classification report
report_df = pd.DataFrame(
    classification_report(y_test, y_test_pred, output_dict=True))

# Overfitting gap
train_acc = model.score(X_train, y_train)
test_acc = model.score(X_test, y_test)
overfit_gap = abs(train_acc - test_acc)

# Confusion Matrix
cm = confusion_matrix(y_test, y_test_pred)
plt.figure(figsize=(5, 4))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.title("Confusion Matrix - Linear SVM (Framing Bias)")
plt.show()

# SHAP values
explainer = shap.LinearExplainer(model, X_train, feature_perturbation="interventional")
shap_values = explainer.shap_values(X_test)
shap.summary_plot(shap_values, X_test, feature_names=X.columns)

# Output metrics
print("\n=== Linear SVM Evaluation - Framing Bias Dataset ===")
print(f"Training Time: {train_time:.4f} seconds")
print(f"Train Accuracy: {train_acc:.4f}")
print(f"Test Accuracy: {test_acc:.4f}")
print(f"Overfitting Gap: {overfit_gap:.4f}")
display(report_df)
```

11.3.3 Random Forest

Figure 43.
Code for Random Forest

```
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, stratify=y, random_state=42)

start_time = time.time()
model = RandomForestClassifier(
    n_estimators=200, random_state=42)
model.fit(X_train, y_train)
train_time = time.time() - start_time

y_train_pred = model.predict(X_train)
y_test_pred = model.predict(X_test)

report_df = pd.DataFrame(
    classification_report(y_test, y_test_pred, output_dict=True))

# Overfitting gap
train_acc = model.score(X_train, y_train)
test_acc = model.score(X_test, y_test)
overfit_gap = abs(train_acc - test_acc)

# Confusion Matrix
cm = confusion_matrix(y_test, y_test_pred)
plt.figure(figsize=(5, 4))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.title("Confusion Matrix - Random Forest (Framing Bias)")
plt.show()

# SHAP
shap_values = explainer.shap_values(X_test)

# If shap_values is a list (older SHAP versions), select class 1
if isinstance(shap_values, list):
    shap_values_to_plot = shap_values[1]
else:
    shap_values_to_plot = shap_values  # Already a NumPy array in new SHAP

# Plot summary
shap.summary_plot(shap_values_to_plot, X_test, feature_names=X.columns)

# Output
print("\nRandom Forest Evaluation - Framing Bias Dataset")
print(f"Training Time: {train_time:.4f} seconds")
print(f"Train Accuracy: {train_acc:.4f}")
print(f"Test Accuracy: {test_acc:.4f}")
print(f"Overfitting Gap: {overfit_gap:.4f}")
display(report_df)
```

11.3.4 XGBoost

Figure 44.
Code for XGBoost

```
from xgboost import XGBClassifier

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, stratify=y, random_state=42)

start_time = time.time()
model = XGBClassifier(
    n_estimators=300,
    learning_rate=0.05,
    max_depth=6,
    random_state=42,
    use_label_encoder=False,
    eval_metric="logloss")
model.fit(X_train, y_train)
train_time = time.time() - start_time

y_train_pred = model.predict(X_train)
y_test_pred = model.predict(X_test)

# Classification report
report_df = pd.DataFrame(
    classification_report(y_test, y_test_pred, output_dict=True)
).transpose()

# Overfitting gap
train_acc = model.score(X_train, y_train)
test_acc = model.score(X_test, y_test)
overfit_gap = abs(train_acc - test_acc)

# Confusion Matrix
cm = confusion_matrix(y_test, y_test_pred)
plt.figure(figsize=(5, 4))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.title("Confusion Matrix - XGBoost (Framing Bias)")
plt.show()

# SHAP
explainer = shap.TreeExplainer(model)
shap_values = explainer.shap_values(X_test)
shap.summary_plot(shap_values, X_test, feature_names=X.columns)

# Output
print("\nXGBoost Evaluation - Framing Bias Dataset")
print(f"Training Time: {train_time:.4f} seconds")
print(f"Train Accuracy: {train_acc:.4f}")
print(f"Test Accuracy: {test_acc:.4f}")
print(f"Overfitting Gap: {overfit_gap:.4f}")
display(report_df)
```

11.3.5 Simple Neural Network

Figure 45.
Code for Simple Neural Network

```
from sklearn.preprocessing import StandardScaler
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout
import time

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, stratify=y, random_state=42)

scaler = StandardScaler()
X_train_s = scaler.fit_transform(X_train)
X_test_s = scaler.transform(X_test)

model = Sequential([
    Dense(32, activation='relu', input_shape=(X_train_s.shape[1],)),
    Dense(16, activation='relu'),
    Dense(1, activation='sigmoid')])

model.compile(
    optimizer='adam',
    loss='binary_crossentropy',
    metrics=['accuracy'])

start_time = time.time()
history = model.fit(
    X_train_s, y_train,
    validation_data=(X_test_s, y_test),
    epochs=20,
    batch_size=32,
    verbose=1)
train_time = time.time() - start_time

train_acc = model.evaluate(X_train_s, y_train, verbose=0)[1]
test_acc = model.evaluate(X_test_s, y_test, verbose=0)[1]
overfit_gap = abs(train_acc - test_acc)

# Predictions
y_pred_probs = model.predict(X_test_s)
y_pred = (y_pred_probs > 0.5).astype(int)

# Classification report
report_df = pd.DataFrame(
    classification_report(y_test, y_pred, output_dict=True))

# Confusion matrix
cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(5,4))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.title("Confusion Matrix - Simple Neural Network (Framing Bias)")
plt.tight_layout()
plt.show()

# Output
print("\n=== Simple Neural Network Evaluation - Framing Bias Dataset ===")
print(f"Training Time: {train_time:.4f}s")
print(f"Train Accuracy: {train_acc:.4f}  |  Test Accuracy: {test_acc:.4f}  |  Overfitting Gap: {overfit_gap:.4f}")
display(report_df)
```

11.3.6 Enhanced Neural Network

Figure 46.

Code for Enhanced Neural Network

```
from tensorflow.keras.layers import BatchNormalization

model = Sequential([
    Dense(128, activation='relu', input_shape=(X_train_s.shape[1],)),
    BatchNormalization(),
    Dropout(0.3),
    Dense(64, activation='relu'),
    BatchNormalization(),
    Dropout(0.3),
    Dense(32, activation='relu'),
    Dropout(0.2),
    Dense(1, activation='sigmoid')])

optimizer = tf.keras.optimizers.Adam(learning_rate=0.001)
model.compile(
    optimizer=optimizer,
    loss='binary_crossentropy',
    metrics=['accuracy'])

start_time = time.time()
history = model.fit(
    X_train_s, y_train,
    validation_data=(X_test_s, y_test),
    epochs=30,
    batch_size=32,
    verbose=1)
train_time = time.time() - start_time

train_acc = model.evaluate(X_train_s, y_train, verbose=0)[1]
test_acc = model.evaluate(X_test_s, y_test, verbose=0)[1]
overfit_gap = abs(train_acc - test_acc)

# Predictions
y_pred_probs = model.predict(X_test_s)
y_pred = (y_pred_probs > 0.5).astype(int)

# Classification report
report_df = pd.DataFrame(
    classification_report(y_test, y_pred, output_dict=True))

# Confusion matrix
cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(5,4))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.title("Confusion Matrix – Enhanced Neural Network (Framing Bias)")
plt.tight_layout()
plt.show()

# Output
print("\n=== Enhanced Neural Network Evaluation – Framing Bias Dataset ===")
print(f"Training Time: {train_time:.4f}s")
print(f"Train Accuracy: {train_acc:.4f}  |  Test Accuracy: {test_acc:.4f}  |  Overfitting Gap: {overfit_gap:.4f}")
display(report_df)
```