pg Cert project proposal

2024/25

Describes the problem area, the aims and objectives of the project, the methodology to solve the problem and the plan towards the proposed solution.

Using Machine Learning/ AI to improve Customer Experience. Enable retailers to forecast customer preferences based on demgraphics

Contents

[1. Project Kernel 2](#_Toc198474262)

[2. Introduction 3](#_Toc198474263)

[2.1 Overview of the project scope and objectives 3](#_Toc198474264)

[2.2 Importance of understanding customer satisfaction in AI-driven shopping experiences 3](#_Toc198474265)

[2.3 Key challenges addressed by project 3](#_Toc198474266)

[3. Project Overview (Heatmap) 4](#_Toc198474267)

[4. Problem Statement 5](#_Toc198474268)

[4.1 Existing gaps in customer satisfaction measurement for AI tools 5](#_Toc198474269)

[4.2 Need for improved data-driven insights into consumer preferences 5](#_Toc198474270)

[4.3 Potential impact on retail and e-commerce strategies 5](#_Toc198474271)

[5. Research Objectives 6](#_Toc198474272)

[5.1 Analyze customer satisfaction trends related to AI tools 6](#_Toc198474273)

[5.2 Identify demographic patterns influencing AI tool adoption 6](#_Toc198474274)

[5.3 Potential impact on retail and e-commerce strategies 6](#_Toc198474275)

[6. Data Acquisition 7](#_Toc198474276)

[6.1 Source datasets from Kaggle 7](#_Toc198474277)

[6.2 Synthetic data 7](#_Toc198474278)

[6.3 Data collection timeframe 7](#_Toc198474279)

[6.4 Justification for dataset selection and merging approach 7](#_Toc198474280)

[7. Data Exploration & Insights 8](#_Toc198474281)

[7.1 Cleaning and encoding categorical data 8](#_Toc198474282)

[7.2 Creation of key features 8](#_Toc198474283)

[7.3 Development of target variables 8](#_Toc198474284)

[7.4 Resampling to enhance dataset size and improve model performance 8](#_Toc198474285)

[8. Data Modelling & Evaluation - Selection of machine learning models 9](#_Toc198474286)

[8.1 K-means algorithm 9](#_Toc198474287)

[8.2 Decision Tree Classifier 9](#_Toc198474288)

[8.3 KNeighbors Classifier 9](#_Toc198474289)

[8.4 XGBoost 9](#_Toc198474290)

[8.5 GaussianNB 9](#_Toc198474291)

[8.6 Logistic Regression 9](#_Toc198474292)

[8.7 Linear Discriminant Analysis 9](#_Toc198474293)

[8.8 Cross-validation results and model performance comparison 9](#_Toc198474294)

[8.9 Selection of Decision Tree Classifier as the preferred model 9](#_Toc198474295)

[9. Expected Impact & Applications 10](#_Toc198474296)

[9.1 Personalised marketing strategies for retailers 10](#_Toc198474297)

[9.2 Improved AI tool adoption insights for businesses 10](#_Toc198474298)

[9.3 Consumer trust and engagement with AI-enhanced shopping experiences 10](#_Toc198474299)

[10. Future Work & Recommendations 11](#_Toc198474300)

[10.1 Addressing class imbalance to refine models 11](#_Toc198474301)

[10.2 Expanding dataset with synthetic data augmentation 11](#_Toc198474302)

[10.3 Broader application of findings to different industries 11](#_Toc198474303)

[11. Conclusion 12](#_Toc198474304)

[11.1 Summary of key findings 12](#_Toc198474305)

[11.2 Final recommendations for stakeholders 12](#_Toc198474306)

[11.3 Next steps for project implementation 12](#_Toc198474307)

# Project Kernel

**INSTRUCTIONS FOR STUDENTS INSTRUCTIONS FOR STAFF**

**(1)** complete boxes A, B **(1)** complete boxes C and D

**(2) Upload the completed form by deadline specified on Moodle**

**A Proposal from:** Rodney Sibanda

**Email: rsiban01@student.bbk.ac.uk**

**Reg No: 14016555**

**B**

**Suggested working title: Customer experience in retail - How I use data analysis to improve customer experience**

**General topic area (e.g. Data Analytics, Databases, Machine Learning, Software Development):**

Data Science; Software Development

**Brief description of problem to be tackled and up to three key references:**

Analysis customers experience data from different sources, identify patterns & anomalies that impact on customer experience and then provide actionable insight/ recommendations for eCommerce platform product owners.

[Online shoppers perceived impact of AI in customer service in 2024](https://www.statista.com/statistics/1484278/online-shopping-experience-ai-customer-service/)

[ACSI - U.S. customer satisfaction with online retail as of 2024](https://www.statista.com/statistics/185760/us-customer-satisfaction-with-e-retail-since-2000/)

[Opinions on the online shopping customer experience 2022](https://www.statista.com/statistics/1377261/attitudes-online-shopping-experience/)

**Level of project difficulty** (from 1 for “appropriate to the degree” to 5 “for very challenging”)**:**

This is 3, integrating/ harmonising the datasets will be the hardest part. I should be able to use Python/ SQL to import the data into a staging environment where I can profile, cleanse and analyse the data using tools I’ve worked with on the course.

If I can manage that, I can talk to my supervisor about stretch targets.

**Resources required**: (list dataset, software and hardware)

Visual Studio and open-source modules/ libraries. Other datasets may be considered after discussion with the supervisor.

**Are these currently available within the School? Yes,** The projects will use open-source tools. I may need to get subscriptions to acquire the datasets

**Is there an outside company involved? Yes, my own limited company, Agile Chameleon.**

**C**

**This project is approved for the PG Cert Applied Data Science YES/NO**

**D** (project tutor) **Status: OK** (accepted)  **RP** (revise)  **RA** (revision accepted)

**Supervisor:**

# Introduction

This section will describe the scope & objectives of the project & provide background information explaining its real-world relevance.

## Overview of the project scope and objectives

The project is to identify the key features/ information a retailer requires to identify a demographic/ class label for a new customer/ prospect.

Classifying that new customer/ prospect will provide insight about products they may be interested in, the type of AI tools they prefer and how frequently they’re likely to purchase products ranged by a specific retailer.

Classification also provides insight into how much they’re likely to spend in each period. This is key in deciding on marketing strategy/ targeting marketing spend.

### Out of scope

This project will not cover: -

* Dynamic Pricing
* Demand forecasting
* Chatbots – customer service
* Visual & Voice Search

These AI Tools enhance customer experience & will be examined in future projects.

## Importance of understanding customer satisfaction in AI-driven shopping experiences

* [AI in Retail: Personalized Shopping](https://businesscasestudies.co.uk/ai-in-retail-personalized-shopping-experiences/)
  + AI analyzes customer behavior to suggest products tailored to individual preferences.
  + Retailers like Amazon use AI-driven recommendation engines to boost engagement and sales.
* Forbes article in folder
  + AI's ability to deliver tailored shopping experiences is a game-changer. AI can analyze vast amounts of customer data—from browsing habits to purchase history—to predict individual preferences with remarkable accuracy. Personalization boosts customer engagement enhances satisfaction and increases sales.
* AI-driven recommendation engines analyse user behaviour, past purchases and preferences to suggest products that customers might like, helping to increase average order value. Platforms like Spotify and Netflix use similar recommendation models, and in e-commerce, these systems drive higher engagement and revenue.
* **• Challenge:** Balancing personalization with over-recommendation, which can overwhelm or annoy customers.
* **• Takeaway For Leaders:** Continuously refine recommendation algorithms with customer feedback and A/B testing to improve their effectiveness.

## Key challenges addressed by project

This project will enable retailers to identify the key information required to classify a new customer/ prospect and show how that information can be leveraged to improve customer engagement and to personalize marketing strategies.

# Project Overview (Heatmap)

**Resources**

**Schedule**

**Scope**

Tasks are marked as follows: on schedule (GREEN), at risk (ORANGE), and behind schedule (RED). Currently, everything is on schedule.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **May 5** | **May 19** | **Jun 2** | **Jun 16** | **Jun 30** | **July 14** | **July 28** |
| **Prepare Project Plan** | **1st pass Data Acquisition** | **Supervisor reviews project proposal** | **Submit Project Proposal** | **Fine tune data acquisition/ data cleansing** | **Fine tune data exploration & data visualisation** | **Supervisor reviews project/ dissertation** |
| **Prepare GitHub Repository** | **1st pass Data Cleansing** | **1st pass Data Visualisation** |  | **Prepare 1st draft of final dissertation** | **Update draft dissertation** |  |
|  | **1st pass Data Exploration** |  |  |  |  |  |
|  | **Update Project Proposal** |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Aug 11** | **Aug 25** |  |  |  |  |  |
| **Update data pipeline & dissertation** | **Submit project and dissertation** |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

# Problem Statement

## Existing gaps in customer satisfaction measurement for AI tools

Over-recommendation & spamming vs hyper personalisation – low conversion/ junk mail. How many times have you flagged a retailers engagement attempts as junk or ignored a retailers recommendations? Use DSG numbers – less than 0.1%

## Need for improved data-driven insights into consumer preferences

Accurate classification will make it easier to predict what a customer may be interested in.

[6 Ways AI Can Improve the Customer Experience](https://www.forbes.com/sites/forbesbooksauthors/2024/04/01/6-ways-ai-can-improve-the-customer-experience/)

[5 ways of enhancing customer experience in retail with AI - BOI (Board of Innovation)](https://www.boardofinnovation.com/blog/5-ways-of-enhancing-customer-experience-in-retail-with-ai/)

[13 ways AI will improve the customer experience in 2025](https://www.zendesk.co.uk/blog/ai-customer-experience/)

## Potential impact on retail and e-commerce strategies

Improved classification will improve customer sentiment – customer engagement and customer retention

# Research Objectives

## ~~Analyze customer satisfaction trends related to AI tools~~

Identify which customers utilise AI tools & which tools are more popular - how AI tools improve engagement & retention

## ~~Identify demographic patterns influencing AI tool adoption~~

How each demographic/ class uses AI Tool

## Potential impact on retail and e-commerce strategies

Will leveraging AI Tools improve customer engagement

[AI in Ecommerce: Top Tools to Boost Customer Satisfaction in 2025](https://wegic.ai/blog/ai-ecommerce-tools.html)

[Artificial Intelligence In E-Commerce: 6 Use Cases And Examples](https://www.cleverence.com/articles/business-blogs/artificial-intelligence-in-e-commerce-6-use-cases-and-examples/)

# Data Acquisition

## Source datasets from Kaggle

## Synthetic data

## Data collection timeframe

## Justification for dataset selection and merging approach

# Data Exploration & Insights

## Cleaning and encoding categorical data

## Creation of key features

## Development of target variables

## Resampling to enhance dataset size and improve model performance

# Data Modelling & Evaluation - Selection of machine learning models

## K-means algorithm

## Decision Tree Classifier

## KNeighbors Classifier

## XGBoost

## GaussianNB

## Logistic Regression

## Linear Discriminant Analysis

## Cross-validation results and model performance comparison

## Selection of Decision Tree Classifier as the preferred model

# Expected Impact & Applications

## Personalised marketing strategies for retailers

## Improved AI tool adoption insights for businesses

## Consumer trust and engagement with AI-enhanced shopping experiences

# Future Work & Recommendations

## Addressing class imbalance to refine models

## Expanding dataset with synthetic data augmentation

## Broader application of findings to different industries

# Conclusion

## Summary of key findings

## Final recommendations for stakeholders

## Next steps for project implementation