

PREDICTING CUSTOMER CHURN IN E-COMMERCE USING MACHINE LEARNING

MSc Research Project

Data Analytics

AHSAN SHAHZAD

Student ID: x23256460

School of Computing

National College of Ireland

Supervisor : SYED MUSLIM JAMEEL

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: AHSAN SHAHZAD
Student ID: x23256460
Programme: MSc in Data Analytics **Year:** 2024-25
Module: MSc Research Project
Supervisor: SYED MUSLIM JAMEEL

Submission Due Date: 15th September 2025

Project Title: Configuration Report

Word Count: 1055

Page Count 6

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: AHSAN SHAHZAD

Date: 15th September 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	☐
Attach a Moodle submission receipt of the online project submission , to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Report

Ahsan Shahzad

X23256460

National College of Ireland

This report goes into great detail about the software, libraries, and configuration settings needed to set up the experimental environment for sentiment analysis and trend forecasting.

Operating System (OS)

- Recommended OS: Windows 10 (64-bit) / Windows 11 (64-bit)
- Alternative OS: Ubuntu 20.04 LTS or macOS Monterey (for Python compatibility)

Tools

- IDE: Jupyter Notebook (via Anaconda) or Google Colab
- Version Control (Optional): Git/GitHub for code management
- Data Source: Kaggle (Customer Churn Prediction – All Features dataset)

Hardware Requirements

- Processor (CPU): Intel Core i5 or higher
- RAM: Minimum 16 GB (32 GB recommended for deep learning models like CNN/LSTM)
- Storage: At least 2 GB free space
- GPU (Optional): NVIDIA GPU with CUDA support for faster CNN/LSTM training

Programming Language

- Language: Python 3.8 or higher

Dataset Configuration

Dataset Name: Customer Churn Prediction – All Features

Source: Kaggle

Size: ~5,880 records, 21 columns

Target Variable: Churn (Yes = 1, No = 0)

Required Python Libraries #

Core Libraries

pandas==1.5.3

numpy==1.23.5 #

Visualization

matplotlib==3.7.1

seaborn==0.12.2 #

Machine Learning scikit-

learn==1.2.2

xgboost==1.7.6

lightgbm==3.3.5 # Deep

Learning tensorflow==2.12.0

keras==2.12.0 # Others

imbalanced-learn==0.10.1 # (if needed for SMOTE)

Model Configurations XGBoost &

LightGBM:

- learning_rate = 0.01 / 0.1
- max_depth = 3 / 5
- n_estimators = 50 / 100
- subsample = 0.8 / 1.0 CNN & LSTM:
- Optimizer: Adam
- Loss: Binary Crossentropy
- Batch Size: 32
- Epochs: 50

Visualization Settings

Tools: matplotlib and seaborn Plot

Types:

- Bar charts (Churn distribution, Contract type vs Churn)
- Density plots (Monthly Charges vs Churn)
- Scatter plots (Tenure vs Monthly Charges)

Style: Use seaborn.set(style="whitegrid") for cleaner visuals

Additional Notes

- Use Google Colab for CNN and LSTM to leverage GPU acceleration.
- Ensure Kaggle dataset is downloaded and placed in the project directory.
- Validate Python version compatibility with libraries listed.
- Always run hyperparameter tuning (GridSearchCV) for XGBoost and LightGBM with 5fold CV for best results.

Table of Contents

Chapter 1: Introduction	
1.1 Introduction	
1.2 Background	
1.3 Research Aim and Objectives	
1.4 Research Question	
1.5 Research Significance	
1.6 Problem Statement	
Chapter 2: Literature Review	
2.1 Introduction	
2.2 Understanding Customer Churn in E-Commerce	
2.3 Traditional Approaches to Churn Prediction	
2.4 Machine Learning Techniques for Churn Prediction	
2.4.1 Logistic Regression	
2.4.2 Random Forest	
2.4.3 Support Vector Machine.....	
2.5 Literature Gap	
Chapter 3: Research Methodology	
3.1 Introduction	
3.2 Knowledge Discovery in Databases (KDD)	
3.2.1 Data Selection	
3.2.2Data Preprocessing	

3.2.3	Data Transformation	
3.2.4	Data Mining	
3.2.5	Evaluation and Interpretation	
Chapter 4: Data Analysis and Findings.....		
4.1	Introduction	
4.2	Exploratory Data Analysis (EDA)	
4.3	Results	
4.3.1	XGBoost	
4.3.2	LightGBM	
4.3.3	Convolutional Neural Network (CNN)	
4.3.4	Long Short-Term Memory (LSTM)	
4.4	Evaluation	
Chapter 5: Conclusion and Recommendation		
5.1	Conclusion	
5.2	Recommendation	
Chapter 6: Future Work.....		
References		

References

- Al Rahib, M.A., Saha, N., Mia, R. and Sattar, A., 2024. Customer data prediction and analysis in e-commerce using machine learning. *Bulletin of Electrical Engineering and Informatics*, 13(4), pp.2624-2633.
- Azevedo, A., 2019. Data mining and knowledge discovery in databases. In *Advanced methodologies and technologies in network architecture, mobile computing, and data analytics* (pp. 502-514). IGI Global Scientific Publishing.
- Berger, P. and Kompan, M., 2019. User modeling for churn prediction in E-commerce. *IEEE Intelligent Systems*, 34(2), pp.44-52.
- De, S. and Prabu, P., 2022. Predicting customer churn: A systematic literature review. *Journal of Discrete Mathematical Sciences and Cryptography*, 25(7), pp.1965-1985.

Hasan, M.S., Siam, M.A., Ahad, M.A., Hossain, M.N., Ridoy, M.H., Rabbi, M.N.S., Hossain, A. and Jakir, T., 2024. Predictive Analytics for Customer Retention: Machine Learning Models to Analyze and Mitigate Churn in E-Commerce Platforms. *Journal of Business and Management Studies*, 6(4), pp.304-320.

Imani, M., Joudaki, M., Beikmohamadi, A. and Arabnia, H.R., 2025. Customer Churn Prediction: A Review of Recent Advances, Trends, and Challenges in Conventional Machine Learning and Deep Learning.

Jahan, I. and Sanam, T.F., 2024. A comprehensive framework for customer retention in Ecommerce using machine learning based on churn prediction, customer segmentation, and recommendation. *Electronic Commerce Research*, pp.1-44.

Li, X. and Li, Z., 2019. A Hybrid Prediction Model for E-Commerce Customer Churn Based on Logistic Regression and Extreme Gradient Boosting Algorithm. *Ingénierie des Systèmes d'Information*, 24(5).

Ljupco Markuseski, L., Zdravkoski Igor, I.Z., Andonovski Miroslav, M.A. and Jovanoska Aleksandra, A.J., 2019, November. Knowledge discovery databases (kdd) process in data mining. In *INTERNATIONAL CONFERENCE PROCEEDING* (pp. 529-539). Faculty of Economic Prilep.

Mammadov, R., 2024. *Customer Churn Prediction*. [Online] Available at: <https://www.kaggle.com/datasets/rashadmammadov/customer-churn-dataset>

Pondel, M., Wuczyński, M., Gryncewicz, W., Łysik, Ł., Hernes, M., Rot, A. and Kozina, A., 2021, July. Deep learning for customer churn prediction in e-commerce decision support. In *Business Information Systems* (pp. 3-12).

Raeisi, S. and Sajedi, H., 2020, October. E-commerce customer churn prediction by gradient boosted trees. In *2020 10th International Conference on Computer and Knowledge Engineering (ICCCKE)* (pp. 055-059). IEEE.

Shaker Reddy, P.C., Sucharitha, Y. and Vivekanand, A., 2024. Customer Churn Prevention For Ecommerce Platforms using Machine Learning-based Business Intelligence. *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)*, 17(5), pp.456-465.

Shikhli, M.S.E. and Hammad, A.M., 2018, June. Data Acquisition Model for Analyzing Schedule Delays Using KDD: Knowledge Discovery and Datamining. In *Proceedings of the 4th International Conference on Frontiers of Educational Technologies* (pp. 67-74).

Shobana, J., Gangadhar, C., Arora, R.K., Renjith, P.N., Bamini, J. and devidas Chincholkar, Y., 2023. E-commerce customer churn prevention using machine learning-based business intelligence strategy. *Measurement: Sensors*, 27, p.100728.

Ulas, B.T., Imer, S., Kalayci, T.A. and Asan, U., 2021, October. Modeling Customer Churn Behavior in E-commerce Using Bayesian Networks. In *Global Joint Conference on Industrial Engineering and Its Application Areas* (pp. 283-300). Cham: Springer International Publishing.

Vanneschi, L., Horn, D.M., Castelli, M. and Popovič, A., 2018. An artificial intelligence system for predicting customer default in e-commerce. *Expert Systems with Applications*, 104, pp.1-21.

Win, T.T. and Bo, K.S., 2020, November. Predicting customer class using customer lifetime value with random forest algorithm. In *2020 International conference on advanced information technologies (ICAIT)* (pp. 236-241). IEEE.

Xia, G. and He, Q., 2018, March. The research of online shopping customer churn prediction based on integrated learning. In *2018 International Conference on Mechanical, Electronic, Control and Automation Engineering (MECAE 2018)* (pp. 259-267). Atlantis Press.

Yayun, Z.H.U.A.N.G., 2018. Research on E-commerce customer churn prediction based on improved value model and XG-boost algorithm. *Management Science and Engineering*, 12(3), pp.51-56.