**Data Science Product Development**

Telecom Customer Churn Analysis

Student Name: Altoglou Athanasios

Student ID: 219082215

WordCount: **2030**

Table of Contents

[1. **Introduction** 3](#_Toc103621494)

[**2.** **Product Design** 3](#_Toc103621495)

[Data source and theme selection and specification 3](#_Toc103621496)

[Application domain/and user’s requirements analysis 3](#_Toc103621497)

[Product functional and non-functional requirements 4](#_Toc103621498)

[Product software architecture design 5](#_Toc103621499)

[Product user case specifications 6](#_Toc103621500)

[**3.** **Product Development** 6](#_Toc103621501)

[The selection of appropriate software tools/platforms and hardware methodologies 6](#_Toc103621502)

[Product development software engineering methodology 7](#_Toc103621503)

[System testing method 8](#_Toc103621504)

[User evaluation plan and methods 8](#_Toc103621505)

[**4.** **Project Management** 9](#_Toc103621506)

[Time Management – Gantt Chart 9](#_Toc103621507)

[Risk assessment on personal information protection and data security/governance 9](#_Toc103621508)

[Quality control on software development 9](#_Toc103621509)

[Basic customer relationship management 9](#_Toc103621510)

[Basic product marketing strategy 10](#_Toc103621511)

[5. Conclusion 10](#_Toc103621512)

[References 11](#_Toc103621513)

Table of Figures

[**Figure 1: Dashboard for telco customer churn analysis** 4](#_Toc103362859)

[**Figure 2: Customer Details Dashboard** 5](#_Toc103362860)

[**Figure 3: Power BI Desktop Software Architecture** 6](#_Toc103362861)

[**Figure 4: Scatterplot of Total Charges versus Monthly Charges** 7](#_Toc103362862)

[**Figure 5: Agile software development cycle** (Zuhdi, 2021) 7](#_Toc103362863)

[**Figure 6: AdaBoost (left) and Smote-ENN (right) evaluation metrics.** 8](#_Toc103362864)

[**Figure 7: Final product** 8](#_Toc103362865)

[**Figure 8: Project Gantt Chart** 9](#_Toc103362866)

# **Introduction**

Over the past years with the advancement of modern technology and the rapid expansion of the market in telecom domain, companies are facing a huge financial loss due to growing competition and loss of potential customers. Customer churn is a major problem in the telecom industry and enterprises are searching ways to prevent this issue as customer retention is the growth pillar for industries. Churn analysis aims to predict customers who are likely to stop using a service or a product during a specific time frame (Gaur A. , Dubey R., 2018)

Working in telecom industry and being able to apply data science methods and algorithms that can help with getting better and useful insights as well as to detect and predict customer churn is extremely important to allow huge financial gains to be achieved within the company. The goal of this report is to critically discuss the task of designing and developing a bespoke data science product for customer churn analysis as a part of an R&D project. The end user of the data science product who have no, or very limited knowledge of these technologies is expected to install and deploy the product in an easy and user-friendly style. Moreover, part of the assignment will be the creation of an interactive dashboard using Power BI Desktop from Microsoft. Finally, the report will look to critically discuss the project management of the data science product.

# **Product Design**

## Data source and theme selection and specification

Product design is extremely important before developing any product, whether it is software or hardware-based product because it interacts with the user and how it is made. Product design deals with the functionality of the product, usability as well as the external appearances. The data product developed will follow the data science road map resulting in the production of a tool (model deployment) which can predict whether a customer will churn or not and a creative dashboard presenting the results in an effective way to ensure that the end-product have all the requirements and functionality for a real-world industry scenario (Candy, 2017).

To start with, the first step was to decide what data and data source was the product going to be built on. A data set found on Kaggle called ‘Telco Customer Churn’ (Kaggle, 2018) which each row represents a customer, and each feature includes customer’s attributes such as demographic information (gender, age range etc), customer account information and services that each of them has signed up for. Therefore, this data set is chosen to build and deploy a machine learning model as well as to create an interactive dashboard for churn analysis following the extract, transform and load (ETL) process.

## Application domain/and user’s requirements analysis

One of the most important factors in the success of product design and development is the capture of the right user requirements. No matter how well structured the other phases are, if someone is unable to get these requirements right, it may lead doing a considerable job of building a data product that was not required (Muhammad Zeeshan Qureshi,Anique Azhar,Qazi Abubakar,Tauseef Ahmad Rana,Ayesha Maqbool, 2021). Focusing on the telecom domain and end-user requirements for the data product, speed, clarity, and accessibility are three main factors which are consider critical when it come to make decisions to secure maximum gains.

## Product functional and non-functional requirements

Understanding the product’s requirements is critical in the stage of planning and designing. Functional requirements are the services that a system should provide to the user and how the system should behave under particular circumstances. On the other hand, non-functional requirements are the functions or services offered by the system (Sommerville, 2016). Planning these requirements can lead for a better-quality product at the end of the project.

Table 1: Product Requirements

|  |  |
| --- | --- |
| **Functional Requirements** | **Non-Functional Requirements** |
| Connect, transform, and model the data | User-friendly and easy to navigate the dashboard |
| The user can navigate the report/dashboard | Enhanced data visualisation with interactive graphs |
| The user can interact with data plots using to different colours for better visual insights | Free and open source with excellent performance, usability, and scalability |
| Deep dive into a specific customer (Customer Details Dashboard) | Responsive software platform allowing audience members to react and progress through the story quickly either from computer or mobile phones. |
| Export the results into a PDF |  |

Calendar

Description automatically generated

**Figure 1: Dashboard for telco customer churn analysis**

Graphical user interface, application

Description automatically generated

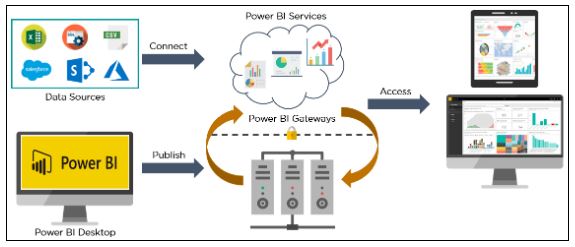
**Figure 2: Customer Details Dashboard**

In figure 2, the user can choose a specific customer to see their details and whether they churn or not using the slicer on top right in customer ID.

## Product software architecture design

Power BI Desktop architecture is a client-server model built on top of Azure. There are many data sources that Power BI can connect to and supply information from such as file types (csv, xml, json etc), databases, azure, online and other services (Hadoop, Vertica etc). Combining the data into a data model, it allows the user to create reports and data visualisations on the dataset. Power BI Gateway is used to keep fresh information by connecting to user’s data sources to get data for analysis (Amrapali Bansal,A. K. Upadhyay, 2017).

Power BI services refer to cloud services which are used to publish reports. As the product requires internet connection, the user can use mobile apps as well and stay connected to their data from anywhere (Biswal, 2022)



**Figure 3: Power BI Desktop Software Architecture**

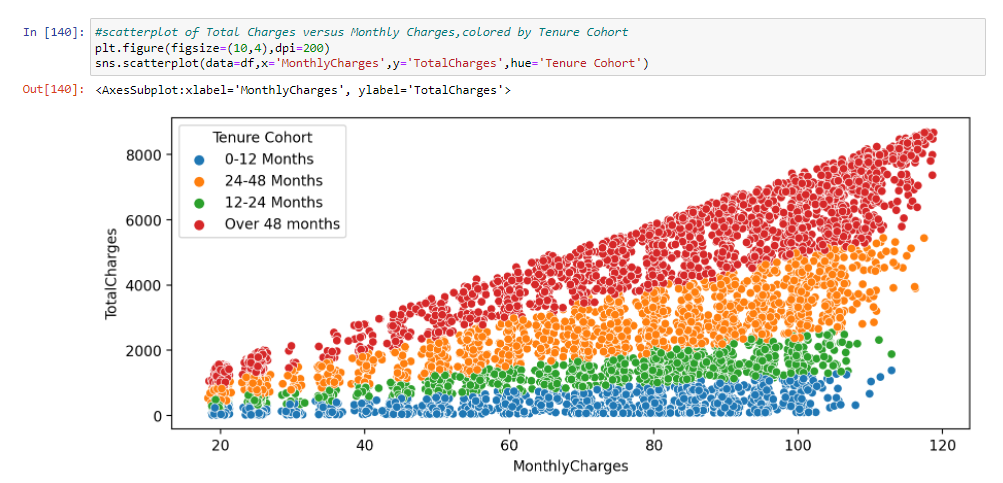
## Product user case specifications

The product is designed to be quickly accessible and easy to use, so people can search information about their customers. Therefore, one main use-case is the data visualisation as it is very crucial for human management in business organizations and services. Power BI provide tools that will let the user visualise key data points using filters and colours to highlight and differentiate important insights. Moreover, users can connect to Power BI through their mobile phones (BYOD/Bring Your Own Device), and they can have a quick access to customer data and the reports.

# **Product Development**

## The selection of appropriate software tools/platforms and hardware methodologies

Part of the data product development is the code and machine learning model deployment to predict customer churning. To produce the code and build the model a bunch of software tools were utilized. Considering these factors, the Jupyter Notebook, a free software tool for interactive computing, and Python programming language were used In ETL phase. Pandas library was used to read the csv file, tidy, transform, manipulate the customer data and for the exploratory data analysis. Then, for data visualisation purposes, matplotlib and seaborn libraries were applied to get better insights through scatterplots, boxplots etc as a good data visualisation can show things that the user did not expect and explain the importance of key data points to visually inclined people (Xuedi Qin, Yuyu Luo, Nan Tang, and Guoliang Li, 2018).

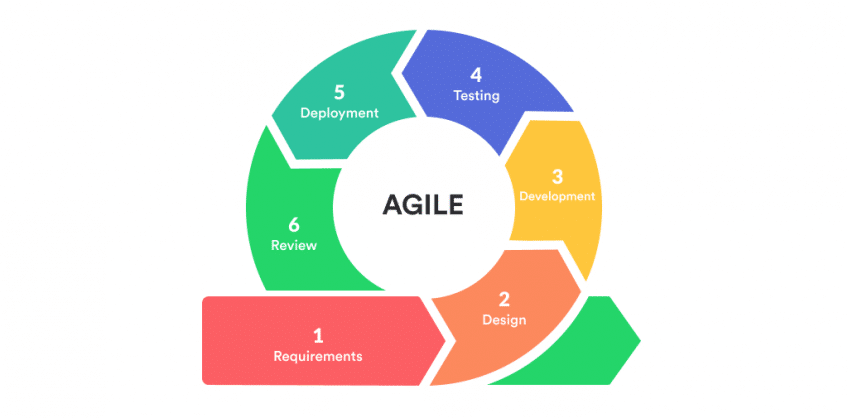


**Figure 4: Scatterplot of Total Charges versus Monthly Charges, coloured by Tenure Cohort**

Building the model, 5 machine learning algorithms compared each other: Decision Trees, Random Forests, Gradient Boost and Adaboost and Smote-ENN (upsampling+ENN) method. After, in order to save and load the final model, python pickle module was imported. To deploy the model with the best performance, an API was created in Jupyter lab to ‘wrap’ it, using the flask library for python that allows someone to create a web application. A web application dashboard was created with HTML in Sublime Text Editor. Apart of these technologies, as it is mentioned before, Power BI Desktop was used to create dashboards and reports.

## Product development software engineering methodology

To create and develop the data product the Agile software engineering methodology was implemented. This method puts a lot of focus on collaboration where people work together to create and respond to changes when necessary, saving time. Therefore, evaluation and planning are the key aspects of agile methodology before and after deploying the product (Zuhdi, 2021).

**Figure 5: Agile software development cycle** (Zuhdi, 2021)

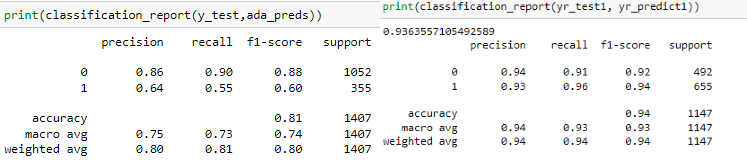
## System testing method

Testing is an essential step in agile methodology because its regular occurrence. Testing process involves that the pipeline of code and the final application are completed without bugs and errors, and it runs in a way which is satisfactory to the user who tests the functional and non-functional requirements.

## User evaluation plan and methods

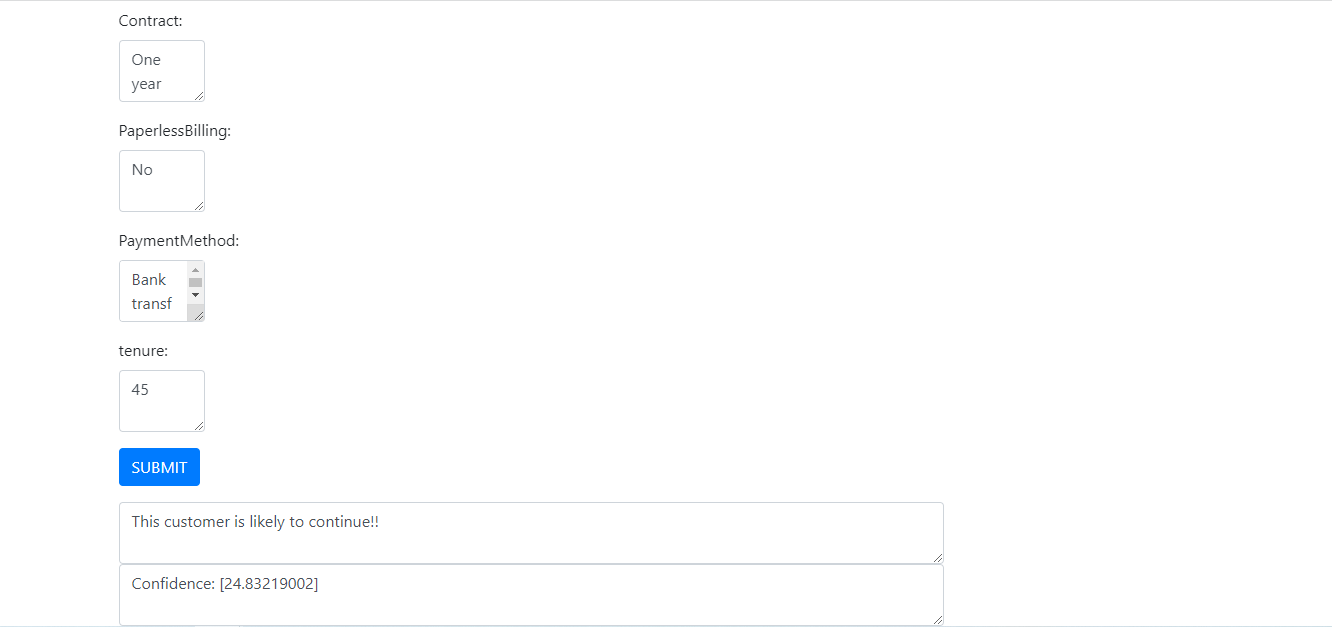
Evaluation was conducted in models’ accuracy and in the final application where the user posts customer’s attributes and an upcoming prediction whether the customer will churn or not takes place as well as the probability of churning.

The following figures breakdown the evaluation metrics of AdaBoost algorithm and Smote-ENN. It can be observed that the smote-ENN method outperforms all the other algorithms with 94% overall accuracy.



**Figure 6: AdaBoost (left) and Smote-ENN (right) evaluation metrics.**

Regarding the application, the API can serve as an interface for GET and POST requests. The user can input and submit customer’s attributes and the application will predict if the customer churned.



**Figure 7: Final product**

Full published product can be found at: https://github.com/Thanasis-Altoglou/Telco-Customer-Churn-Analysis

# **Project Management**

## Time Management – Gantt Chart

**Figure 8: Project Gantt Chart**

As part of the project, a Gantt chart was used to visualise the project plan with building blocks for each activity against time. Gantt charts helps groups to properly allocate resources, monitor the schedule of the project and make changes or adjustments as necessary. Moreover, this chart makes it easier to create more complicated tasks and work around deadlines (Robbin, 2019).

## Risk assessment on personal information protection and data security/governance

Risk evaluation is an extremely important aspect when working on any data product, especially in churn analysis because of the customers’ personal data and information. If customer data is not protected correctly in the process of use, it will directly risk their privacy and the security of data (Zhang, 2018). Since the security of data is a must requirement because of GDPR, it is vital that data should be handled in an efficient way. To realise the protection and security of the data, a large number of professional security technologies are needed. However, for the specific project, the data set is publicly available on Kaggle, so this is not a concern.

## Quality control on software development

Quality control on the software development project is very important because it ensures that the end data product is running and can be released without breaking or having bugs and errors and provide a sufficient and up to the mark experience. In the agile software development lifecycle, developers focus a lot in the testing phase where they can find out if their programming and code works or the final product meets the customer requirements.

## Basic customer relationship management

Customer relationship management can be defined as the combination of operations, strategies, and technologies that an organisation uses to manage customers interactions with current and potential customers. Its goal is to improve customer service relationship, drive sale growth and assist in customer retention (Subashini,Apte,Rajesh,Praveen,Rai, 2017). Through CRM approach, companies can learn more about their customers and cater to their needs. For example, Salesforce is a software tool that let someone store customer data and information, recognise sales opportunities and deal with marketing campaigns.

## Basic product marketing strategy

Customer retention is the growth pillar for companies to make profit, so they need to develop their strategies to prevent a customer from leaving. Telecom companies can implement some techniques using their CRM system to retain their customers. They can target potentially churners with exclusive offers based on their purchase history as well as to automate emails via trigger-based events which have higher open rates.

# Conclusion

Customer churn is one of the biggest issues in telecommunication industry as churn prediction and prevention has become an essential task for customer relationship management. Because of the high competition and the products and offerings comparison, customers are less loyal.

The goal of this project was to take a closer look at the telecom domain and customer data and create a data product to predict whether a customer will churn or not. Smote-ENN method outperform the other ML models with 94% overall accuracy. However, every data science team should have a full stack web developer to creative more attractive dashboards for the final application.

# References

Amrapali Bansal,A. K. Upadhyay, 2017. Microsoft Power BI. *International Journal of Soft Computing and Engineering (IJSCE),* 7(3).

Biswal, A., 2022. *Simplilearn.* [Online]   
Available at: https://www.simplilearn.com/tutorials/power-bi-tutorial/what-is-power-bi

Candy, F., 2017. *The Data Science Handbook John Wiley & Sons.* NJ ed. s.l.:Hoboken.

Gaur A. , Dubey R., 2018. Predicting Customer Churn Prediction In Telecom Sector Using Various Machine Learning Techniques.

Kaggle, 2018. *Kaggle.* [Online]   
Available at: https://www.kaggle.com/datasets/blastchar/telco-customer-churn

Muhammad Zeeshan Qureshi,Anique Azhar,Qazi Abubakar,Tauseef Ahmad Rana,Ayesha Maqbool, 2021. Capturing Users Requirements Using a Data Mining. *International Conference on Communication Technologies.*

Robbin, T. R., 2019. A Multipart Project Planning and Tracking Exercise. *Decision Sciences Journal of Innovative Education,* 17(2).

Sommerville, I., 2016. *Software Engineering.* 10th ed. s.l.: Pearson Education.

Subashini,Apte,Rajesh,Praveen,Rai, 2017. Churn Prediction in Customer Relationship Management via GMDH-Based Multiple Classifiers Ensemble. *Advances in Computational Sciences and Technology,* 10(8).

Xuedi Qin, Yuyu Luo, Nan Tang, and Guoliang Li, 2018. DEEPEYE: An Automatic Big Data Visualization Framework. *IG DATA MINING AND ANALYTICS,* Volume 1.

Zhang, D., 2018. Big Data Security and Privacy Protection. *8th International Conference on Management and Computer Science (ICMCS 2018),* Volume 77.

Zuhdi, M., 2021. *Medium.* [Online]   
Available at: https://mzuhdiz.medium.com/agile-software-development-a-game-changer-of-software-development-process-d52933246c28