Architectural Decisions Document

Project: Predict customer churn

J.Kuntze - March 24, 2020

This document presents the decisions made to create a system to predict customer churn at a telecommunications company.

This document is based on the following articles written by Romeo Kienzler:

- The Lightweight IBM Cloud Garage Method for Data Science https://developer.ibm.com/technologies/artificial-intelligence/articles/the-lightweight-ibm-cloud-garage-method-for-data-science
- Architectural decisions guidelines https://developer.ibm.com/technologies/artificial-intelligence/articles/data-science-architectural-decisions-guidelines

1.1 Data Source

1.1.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

- Data stored in a CSV file will be used as data source.

1.1.2 Justification

Please justify your technology choices here.

- The data set is small enough to be stored locally in a CSV file.

1.2 Enterprise Data

1.2.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

- No enterprise data will be required in this project.

1.2.2 Justification

Please justify your technology choices here.

- Not required because the data is stored locally.

1.3 Streaming analytics

1.3.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

- Streaming will not be required; analyses will be performed in batches.

1.3.2 Justification

Please justify your technology choices here.

- Streaming will not be required; analyses will be performed in batches.

1.4 Data Integration

1.4.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

- Additional tools for data integration (such as Apache Spark or SQL databases) are not required.

1.4.2 Justification

Please justify your technology choices here.

- Additional tools for data integration are not required due to the limited amount of data.

1.5 Data Repository

1.5.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

- Data will be stored locally in a CSV file.

1.5.2 Justification

Please justify your technology choices here.

- Advanced data repository options are not required due to the limited amount of data.

1.6 Discovery and Exploration

1.6.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

- Jupyter Notebook, Python, pandas and Matplotlib.

1.6.2 Justification

Please justify your technology choices here.

- Python and Matplotlib are sufficient to explore the data. Jupyter Notebook and Pandas make it easier to accomplish this task and document the findings.

1.7 Actionable Insights

1.7.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

- Jupyter Notebook, Python, pandas, scikit-learn and Keras.

1.7.2 Justification

Please justify your technology choices here.

- I am familiar with the technolopgy listed above. Limited amount data makes it feasible to complete the project using it.

1.8 Applications / Data Products

1.8.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

- The data products will be a web page containing this project report and the architectural decision document, and a presentation. Jupyter Notebooks, slidify in R, HTML, CSS will be used.

1.8.2 Justification

Please justify your technology choices here.

- An analysis report and a presentation are sufficient to describe this project. If the model were to be used in

production, it could be part of a scheduled report to assist the customer services department to prioritize customers at risk of terminating their contracts. A dashboard could be provided to management on the percentage of customers at risk of terminating their contracts.

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice

Please describe what technology you have defined here. Please justify below, why. In case this component is not needed justify below.

- For this project, only the data scientist will have access to the data. The data will be stored locally.

1.9.2 Justification

Please justify your technology choices here.

- Limited amount of data and open source nature of the dataset do not require additional steps to limit access and manage security of the dataset and data product.