PREDICTIVE ANALYTICS CASE STUDIES ANALYSES

This document consists of two sections that contain the following information:
1.   Analyses of predictive analytics case studies.
2.   Analysis of the results produced by the case study analysis.

1.   Case Study Analyses

Case Study 1: Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company case study[1]

Initiation

Pro requirement 1.1: The goal of the analytics project is the creation of a 'models able to determine customer loyalty'.
Gen requirement 1.1: The goal of the analytics project must be defined.

Pro requirement 1.2: Within the predictive analytics project 'Prediction can be directed at customer loyalty to identify both customers who have great loyalty to their preservation as well as customers with inten- tions to change to the competitors.'
Gen requirement 1.2: The object to which the analytics is directed must be defined.

Acquisition

Pro requirement 1.3: The following data sources must be used within the analytics project 'Call log, SMS message log, MMS Multimedia Message log MMS multimedia mes- saging log, DATA internet data usage log, Mon fee log, Vou recharge log, Mon monthly log information, web metadata information, EGGSK tab In roaming'.
Gen requirement 1.3: The sources of data that will be utilized in the analytics project must be defined.

Pro requirement 1.4: The following feature will be used in used to train the predictive model:
'• Segmentation Features T, F, M (3 features)
total of calls and Internet duration in a certain period of time (Fig 2). Frequency (F): use services frequently within a certain period (Fig. 3). Monetary (M): The money spent during a certain period (Fig. 4).
Classification Features (220 features)
• Individual Behavioral Features
Individual behavior can be defined as how an individual behaves with services. For example: Calls duration per day: calls duration per day for each GSM.
Duration per day: calls and sessions duration per day for each GSM (Figs. 5, 6, 7). Entropy of duration
High entropy means the data has high variance and thus contains a lot of informa- tion and/or noise.
Daily outgoing calls: for each GSM the daily outgoing calls.
Calls incoming daily night: for each GSM the daily outgoing calls at night (Fig. 8). SMS received daily at work time, . . . About (200 features).
• Social behavior features
Is behavior among two or more organisms within the same species, and encompasses any behavior in which one member affects the other. This is due to an interaction among those members.
spatial and navigation features'
Gen requirement 1.4: The features that will be used to train the predictive model must be defined.

Pro requirement 1.5: The analytics project must utilize the following feature selection methods:
'Filter model
Wrapper model
Embedded model'
Gen requirement 1.5: The feature selection methods utilized in the analytics project must be defined.

Pro requirement 1.6: Within the analytics project the 'data was divided into two sub-parts: the training group and the test
group by 70/30, respectively.'
Gen requirement 1.6: The splitting of the data set must be defined for the analytics project.

Pro requirement 1.7: The predictive analytics project must use 'HDP framework was custom-installed' including 'Hadoop HDFS for data storage', 'Yarn for resource management, Ambari for system monitoring', 'Hive is an ETL and data warehouse tool on top of the Hadoop ecosystem and used for processing structured and semi-structured data' for the acquisition and storage of data.
Gen requirements 1.7: The tools and software utilized for the acquiring and storing the data must be defined along with their utility within the project.

Analysis
Pro requirement 1.8: The following software will be used for the creation of the predictive model: ''Spark implementation engine for data processing', 'Zeppelin as a development user interface'.
Gen requirement 1.8: The tools and software utilized for the creating and testing the predictive model must be defined along with their utility within the project.

Pro requirement 1.9: The analytics project will test five algorithms they are as follows 'Multilayer perceptron classifier', 'Gradient Boosted Tree', 'Decision Tree Classifier', 'Random Forest Classifier'.
Gen requirement 1.9: The predictive model(s)/ algorithm(s) that must be used within the analytics project must be defined.

Pro requirement 1.10: The following performance measurements must be used for the within the analytics project:
'True Positive (TP) is expressed as an example when the prediction is yes (the cus- tomer has loyalty to the company), and the truth has loyalty to the company.
True negative (TN): When the prediction is no (no customer loyalty to the com- pany), in fact the customer has no loyalty to the company.
False Positive (FP): The prediction is yes (the customer has loyalty to the company), but the customer left the company is also known as "Type 1 error".
D False negative (FN): When the prediction is not (the customer has no loyalty to the company), but the customer has loyalty and did not actually leave the company. Also known as "Type 2 error".'
Gen requirement 1.10: The performance measures used to evaluate any created models must be defined alongside the acceptable ranges for the values of these measures.
Presentation

Pro requirement 1.11: The following outputs must be provided by the predictive model 'identify the causes of loyalty and to identify the influential features at each level of loyalty ', the loyalty classifications are as follows

'Very high value customers (greater loyalty) These are the customers who make the highest profit for the operator. Without them the operator will lose its market share and competitive advantage. These customers are given appropriate care and atten- tion from the operator. High value customers (great loyalty) These are the customers who make the high- est profit for the operator. Without them, the operator will lose its market share and competitive advantage. These customers are given appropriate care and attention from the operator. Medium value customers (average loyalty) These are the customers who make medium profitability.

Low value customers (little loyalty) These are the clients who make very little profit. Customer churn from the company (very little loyalty).'

'build an accurate predictive model for classify- ing new users by loyalty', 'assist in decision-making in building marketing presentations for each category thus increasing the company's profit'.

Gen requirement 1.11: The output of the model must be defined, alongside the excepted utility of the models' outputs.

Case Study 2: Operating Data-driven Predictive Analytics for Tele-diagnosis of Refrigeration Systems: A Case Study [2]

Initiation

Pro requirement 1.1: The analytics project is having 'Predicative analytics is applied to the case of a refrigeration service provider'.

Pro requirement 1.2: The predictive model must be aimed at 'malfunctioning refrigeration systems.

Acquisition

Pro requirement 1.3: The source of data within the analytics project will be 'iPro controller technology to collect data from refrigeration operations.

Pro requirement 1.4: The features used to train the model will be 'performance metrics based on the trends from three available sensors.

Pro requirement 1.5: Not mentioned.

Pro requirement 1.6: Not mentioned.

Pro requirement 1.7: The analytics project must utilize the 'iPro controller technology to collect data from refrigeration'.

Analysis

Pro requirement 1.8: The analytics project must utilize the 'vapor compression refrigeration (VCR) thermodynamics can be combined with the kinetic and Bernoulli theorems to develop a model for describing the functionality of a refrigeration system'.

Pro requirement 1.9: This analytics project 'Weka machine learning software to cluster different failure types and performance metrics based on the trends from three available sensors'.

Pro requirement 1.10: Not defined.

Presentation
Pro requirement 1.11: The analytics project must create a model that 'predict when sensors were trending towards a future failure'.

Pro requirement 2.1: The analytics project must 'apply data-driven predictive analytics to tele-diagnosis of refrigeration system by designing a robust smart mobile monitoring application'
Gen requirement 2.1: The analytics project must specify the medium (mobile application, analytics dashboard, etc..) by which the predictive analytics model is utilized.

Case study 3: A Case Study of Sri Lanka Oil Price Fluctuations and Its Influencing Factors using Predictive Analytics [3]
Initiation
Pro requirement 1.1: The analytics project must undertake the creation of a data warehouse that enable the use 'analyze the data patterns and forecast future sales and fluctuations based on predictive analytics'.

Pro requirement 1.2: The analytics project must be concerned with the 'comprehensive analysis of oil price fluctuation and supply variation using predictive analysis'.

Acquisition
Pro requirement 1.3: The analytics project can utilize the following data sources 'SAP system, Parliament web site and central bank economic reports'.

Pro requirement 1.4: Not defined.

Pro requirement 1.5: Not defined.

Pro requirement 1.6: Not defined.

Pro requirement 1.7: Within the analytics the 'data warehouse design is implemented under Microsoft SQL Server'.

Analysis
Pro requirement 1.8: Not defined.
Pro requirement 1.9: Not defined.

Pro requirement 1.10: Not defined.

Presentation
Pro requirement 1.11: The model must predict the 'price and quantity variation' of different types of fuels.

Pro requirement 2.1: The model must create 'generate reports in oil related predictions is feasible and efficient'.

Case study 4: Predictive Analytics of Donors in crowdfunding platforms: A case study on Donorschoose.org [4]

Initiation

Pro requirement 1.1: The goal of the analytics project is to 'address the problem of retaining the existing donors by applying the machine learning models to analyze the behavior of the donors and filter the most promising projects'.

Pro requirement 1.2: The analytics project will 'apply machine learning algorithms, and we analyze each feature of the Donorschoose dataset' which represents the 'DonorsChoose.org'.

Acquisition

Pro requirement 1.3: The source of data utilized with the analytics project must be the 'https://www.donorschoose.org/data' webpage.

Pro requirement 1.4: 'The input features for the machine learning algorithms are: Project title, Project short essay, Project categories, Project subcategories, Project submitted Date Time, Project grade category, number of projects proposed by teacher, Donor name, Donor State, Donation time details, School name, School metro type, School state, School Zip, School county, School city, Teacher prefix, Teacher first project proposed time, Resource items, Resource quantity, Resource unit price, Resource vendor name, Donor details etc.'

Pro requirement 1.5: The analytics project must utilize 'Principal component Analysis (PCA) a dimensionality reduction technique to reduce the data features after pre-processing'

Pro requirement 1.6: Not specified.

Pro requirement 1.7: Not specified (Dataset was readily available).

Pro requirement 4.1: The analytics project must vectorize the 'Project Short essay' feature into a 'Bag Of Words model', 'Term frequency and Inverse Document Frequency'.
Gen requirement 4.1: The analytics project must define transformation(s) done to the data before being inputted into the predictive model.

Analysis

Pro requirement 1.8: The analytics project will test 'K-Nearest Neighbor (K-NN)', 'Naïve Bayes classifier', ' Random forests', 'Neural networks'.

Pro requirement 4.2: The following hyperparameters must be tested: KNN – 49 and 45, Naïve Bay – 0.5 and 0.1, Logistic regression – 0.001 and 0.1.
Gen requirement 4.2: The hyper parameters for that will set for the machine learning models must be specified.

Pro requirement 1.9: The analytics project is 'project is completely developed using Jupyter which is a part of Anaconda distribution'.

Pro requirement 1.10: The analytics models created within the project must be evaluated using 'ROC-AUC'.

Presentation

Pro requirement 1.11: The result of the analytics project '10 most important factors for a project to be picked by a donor'.

Pro requirement 2.1: Although not explicitly mentioned it is possible to deduce based on the results of the predictive analytics models that the result is a report showing 10 most important features for a school project to get selected.

Case study 5: Automation in Financial Reporting by using Predictive Analytics in SAP Analytics Cloud for Gold Mining Industry: a Case Study [5]

Pro requirement 1.1: The goal of the analytics project is 'implementation of the predictive analysis model ', 'in the finance module'.

Pro requirement 1.2: The predictive analytics must be focused on the 'Capex finance report'.

Acquisition
Pro requirement 1.3: The analytics project must utilize 'finance report data'.

Pro requirement 1.4: Not specified.

Pro requirement 1.5: Features are selected using the '"Run a Smart Discovery" is used to explore data using machine learning algorithms to find key influencers'.

Pro requirement 1.6: Not specified.

Pro requirement 1.7: The analytics project must utilize 'SAP BW' for acquiring data.

Pro requirement 4.2: Not specified.

Analysis
Pro requirement 1.8: The algorithms included in the 'SAP Analytics Cloud feature, namely Smart Discovery which functions to provide predictive analysis of data using machine learning'.

Pro requirement 1.9: The analytics project must 'SAP Analytics Cloud is a solution in the form of Machine Learning and artificial intelligence provided by SAP'.

Pro requirement 1.10: The evaluation of the results of the predictive model must be given in the form of a 'report on the results of black box testing from the results of the SAP Analytics Cloud'.

Presentation
Pro requirement 1.11: The model must provide a 'predictive analysis display on the Capex finance report for analysis and predictions for the following year'

Pro requirement 2.1: The predictive analytics model will be utilized by an 'ERP dashboard application system is needed to support monitoring carried out by the Board of Directors in each division, especially the Finance division'.

2. Analysis of Case Study Analyses

Shown below are all the generic requirements that were defined through the analysis of case studies.

Initiation
Gen requirement 1.1: The goal of the analytics project must be defined.

Gen requirement 1.2: The object to which the analytics is directed must be defined.

Acquisition
Gen requirement 1.3: The sources of data that will be utilized in the analytics project must be defined.

Gen requirement 1.4: The features that will be used to train the predictive model must be defined.

Gen requirement 1.5: The feature selection methods utilized in the analytics project must be defined.

Gen requirement 1.6: The splitting of the data set must be defined for the analytics project.

Gen requirements 1.7: The tools and software utilized for the acquiring and storing the data must be defined along with their utility within the project.

Gen requirement 4.1: The analytics project must define all preprocessing done to the data before being inputted into the predictive model.

Analysis
Gen requirement 1.8: The predictive model(s)/ algorithm(s) that must be used within the analytics project must be defined.

Gen requirement 4.2: The hyper parameters for that will set for the predictive models must be specified.

Gen requirements 1.9: The tools and software utilized for the development of the predictive model must be defined along with their utility within the project.

Gen requirement 1.10: The performance measures used to evaluate any created models must be defined alongside the acceptable ranges for the values of these measures.

Presentation
Gen requirement 1.11: The output of the predictive analytics model must be defined, alongside the excepted utility of the models' outputs.

Gen requirement 2.1: The analytics project must specify the medium (mobile application, analytics dashboard, or report) by which the predictive analytics model is utilized.

Gen requirement 5.1: The compatibility that relates to medium to which the result of predictive analytics model must be specified where applicable.

Table 1 shows the definition and validation of generic requirements (case study analysis is shown in the appendix) for predictive analytics projects. In the table Y stands for Yes meaning that the generic requirement could be defined for a CS (Case Study) and N means the generic requirement could not be defined for the CS.

Gen(Prd)	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	1.11	2.1	4.1	4.2
CS1	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y			
CS2	Y	Y	Y	Y	N	N	Y	Y	Y	N	Y	Y		
CS3	Y	Y	Y	N	N	N	Y	N	N	N	Y	Y		
CS4	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y
CS5	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	N	N

Table 1: Validation of generic requirements for predictive analytics.

The results of the validation show that Gen requirements 1.6 was not validated, meaning that the all the numeric labels for requirements proceeding Gen requirement 1.6 will be decremented by one accordingly. Meaning the labels for gen requirements will be changed from 1.7 to 1.6, 1.8 to 1.7, 1.9 to 1.8, 1.10 to 1.9, and 1.11 to 1.10.

The finalized generic requirements templates is as follows:

Initiation phase:
1. Gen(prd) requirement 1.1: The goal of the analytics project must be defined.
2. Gen(prd) requirement 1.2: The object to which the analytics is directed must be defined.

Acquisition phase:
3. Gen(prd) requirement 1.3: The sources of data that will be utilized in the analytics project must be defined.
4. Gen(prd) requirement 1.4: The features that will be used to train the model must be defined.
5. Gen(prd) requirement 1.5: The feature selection methods utilized in the analytics project must be defined.
6. Gen(prd) requirement 1.6: The tools and software utilized for acquiring and storing the data must be defined along with their utility within the project.
7. Gen(prd) requirement 4.1: The analytics project must define all pre-processing done to the data before being inputted into the predictive model.

Analysis phase:
8. Gen(prd) requirement 1.7: The model(s)/ algorithm(s) that must be used within the analytics project must be defined.
9. Gen(prd) requirement 1.8: The tools and software utilized for the development of the model must be defined along with their utility within the project.
10. Gen(prd) requirement 1.9: The performance measures used to evaluate any created models must be defined alongside the acceptable ranges for the values of these measures.

Presentation phase:
11. Gen(prd) requirement 1.10: The output(s) of the model must be defined, alongside the excepted utility of the models' outputs.
12. Gen(prd) requirement 2.1: The analytics project must specify the medium (mobile application, analytics dashboard, or report) by which the created analytics model is utilized.

References

[1] W. N. Wassouf, R. Alkhatib, K. Salloum, and S. Balloul, "Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company case study", doi: 10.1186/s40537-020-00290-0.

[2] T. Lu, J. Du and R. J. Jiao, "Operating Data-driven Predictive Analytics for Tele-diagnosis of Refrigeration Systems: A Case Study," 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Bangkok, Thailand, 2018, pp. 1647-1651, doi: 10.1109/IEEM.2018.8607436.

[3] D. S. A. Kandawala, M. A. M. Mansoor, R. T. Ramanayake, D. M. Wanniarachchi, K. G. L. Bogahawatte and P. P. G. D. Asanka, "A case study of Sri Lanka oil price fluctuations and its influencing factors using predictive analytics," 2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS), Galle, Sri Lanka, 2016, pp. 1-6, doi: 10.1109/ICIAFS.2016.7946518.

[4] J. V. Dirisam, D. Bein and A. Verma, "Predictive Analytics of Donors in crowdfunding platforms: A case study on Donorschoose.org," 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), NV, USA, 2021, pp. 0834-0838, doi: 10.1109/CCWC51732.2021.9376094.

[5] S. Nararya, M. Saputra and W. Puspitasari, "Automation in Financial Reporting by using Predictive Analytics in SAP Analytics Cloud for Gold Mining Industry: a Case Study," 2021 International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, 2021, pp. 1-6, doi: 10.1109/ICISS53185.2021.9532504.