BELGIUM CAMPUS iTversity It's the way we're wired	**Assessment:**	**Project**
	Subject:	**Business Intelligence 381**
	Total:	**100**

Recommender Model for Customer Service Eligibility

Introduction

LangaSat is a company offering a new satellite internet service on contract terms, and eligibility is currently determined solely by annual salary. However, the company believes that additional factors may influence eligibility. Your task is to develop an intelligent recommender model capable of accurately identifying customers that have good or bad credit risks and are therefore eligible or not for the service based on various variables other than just the salary.

Besides the recommender model, you may come up with your own data mining goals (classification, market basket analysis, association rule mining, clustering, trend analysis, prediction, etc.) depending on the nature of data. However, these should be approved by your lecturer who is the project leader and supervisor during the first week before you can proceed with the project. **Proceeding with the project without consulting your supervisor may result in severe penalties.**

Outline

You are given a dataset "***CustData2.csv***" that contains customer information such as job titles, departments, salaries, year of birth, marital status, city of residence, years of residence, level of education, occupation, and household size. Currently, only customers earning more than R50 000 qualify for the service contract but over the years, it has been suggested that other factors other than the annual salary have a bearing on the credit risk and eligibility of a customer to qualify for the service contract. You are therefore required to use the given dataset to build a classification model that accurately identifies customers eligible for the service contract based on other variables besides the salary.

The project should follow the Cross-Industry Process for Data Mining Methodology (CRISP-DM) stages shown on Figure 1 below to plan and implement the data mining and business intelligence project. CRISP-DM (Chapman et al., 1999) was initially created as an open standard for data mining processes across industries but has since become the most common methodology for data mining, analytics, and data science projects (Schröer et al., 2021). Therefore, understanding the CRISP-DM framework is very important before beginning this project.

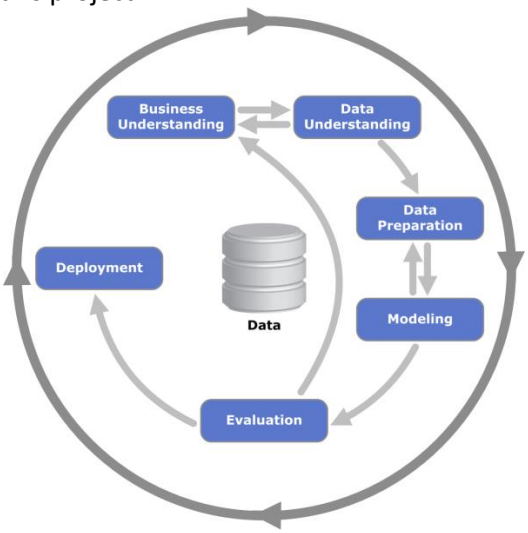

Figure 1: CRISP-DM (Wikimedia Commons, 2012)

BELGIUM CAMPUS iTversity It's the way we're wired	**Assessment:**	**Project**
	Subject:	**Business Intelligence 381**
	Total:	**100**

The project outline underscores the use of CRISP-DM as a structured approach to guide you through the data science project, emphasizing its importance in the context of real-world data analysis and modelling. You may need to gather extra information on related projects and scenarios to establish a strong business case for the project.

Project Steps

The entire project consists of the following steps which are then broken into milestones that should be submitted as assignments on the given due dates.

1. **Business Understanding:**
   - Clearly define the business problem and objectives.
   - Identify stakeholders and their requirements.
   - Determine the success criteria for the recommendation system.
   - Recognize the importance of CRISP-DM as the methodology to guide the project.
2. **Data Understanding:**
   - Explore the provided dataset, including variables like job titles, departments, salaries, etc.
   - Document data quality problems, missing values, duplicates, and outliers.
   - Perform preliminary data visualizations (e.g., correlation matrices, heatmaps, or pair plots) and analysis to understand the significance of variables and gain initial insights into the data.
   - Create preliminary dashboards to visualize data distributions, correlations, and initial insights to identify patterns and relationships in the data before moving on to the modelling phase.
3. **Data Preparation:**
   - Clean and preprocess the data to address quality issues (e.g., impute missing values, remove duplicates, handle outliers).
   - Transform and encode categorical variables as needed.
   - Discretize or scale numerical variables if necessary.
   - Split the data into training and testing sets for model evaluation.
4. **Modelling:**
   - Select a classification algorithm(s) (e.g., logistic regression, decision tree, random forest) to build the recommendation system. Justify the choice of algorithm based on the data characteristics.
   - Use the existing "annual salary" column as initial class labels (dependent variable) but you are not limited to specific variables for predictor (independent) variables.
   - Explore and identify significant variables to be used for classification based on your analysis.
   - Justify your choice of predictor variables by employing methods to evaluate the interestingness, importance, and relevance of these variables.
   - Train and fine-tune the classification model using the training data.
   - Visually represent the model's performance metrics, confusion matrices, and other evaluation results to facilitate interpretation.
   - Document the model selection process and parameter tuning, including the rationale behind predictor variable selection.
5. **Model Evaluation:**
   - Evaluate the classification model's performance using appropriate metrics (e.g., accuracy, precision, recall, F1-score).
   - Compare the model's performance against the baseline (using only salary for eligibility).

- Visualize and interpret the results.

6. **Deployment:**
   - Discuss deployment options for the recommendation system.
   - Develop and document a plan for deploying the model in a real-world scenario.
   - Deploy the final model in a user-friendly interface.
   - Address considerations such as data input, post-deployment monitoring, and updating the model.
   - Consider ethical and privacy implications of deploying the model, as guided by CRISP-DM.
7. **Final Report:**
   - Summarize the entire project, including the business problem, data exploration, data preparation, modelling, evaluation, and deployment plan highlighting the use of CRISP-DM as the methodology guiding the project.
   - Provide insights and recommendations based on the model's findings.
   - Include visualizations, code snippets, and explanations throughout the report.
   - Submit the project report along with any code used for analysis and modelling.

Deliverables:

- Project report (in PDF or document format) detailing all project steps, key findings, ethical implications and recommendations for improvement.
- Code (R Markdown and project files) used for data analysis and modelling.
- Power BI project file if Power BI was used at any stage during the project.
- Presentation of the model deployment in a user-friendly interface.

Grading Criteria:

Criteria	Weight
Understanding of the business problem	10%
Data exploration and quality assessment	15%
Data preprocessing and transformation	15%
Model selection and training	20%
Model evaluation and comparison	15%
Deployment plan and considerations	15%
Documentation and presentation	10%

Additional Information

- This is a group project, but the group may not exceed four people.
- All work must be original. Copying work from another group or any other sources will not be tolerated.
- Includes names of all group members on the Cover Page.
- Submit your project milestones electronically on Moodle (BC Connect) before the due dates.
- All writing must be correctly cited and referenced.
- **Plagiarism is a serious offence**. Belgium Campus uses software that can scan for plagiarism and a student caught doing this will get 0 for this assignment.
- No mark will be awarded if the assignment is not uploaded via BC Connect.
- Late assignments will not be accepted; missing the deadline is an automatic 0.