Introduction:

Implement the CRISP-DM methodology where a comprehensive review of the customer's requirements supports the creation of a business objective outlining Items such as the expected level of model performance and return on investment.
Create a data mining objective which will guide the subsequent work in the data understanding, preparation and modelling steps and the final evaluation and selection of a classification model or models
This project will investigates data mining of demographic data in order to create one or more classification models capable of accurately identifying individuals whose salary exceeds a specified value.

The data used in this project contain information on individuals such as age, level of education and current employment type. For this purpose, the classification model will be used to select candidates for a new service offered by the sponsor of the project targeting individuals with salaries exceeding R50 000

Provide descriptions of the predictive significance of each attribute, interesting or useful patterns found and any transformations applied to the data

The project is divided into seven stages , where at each stage some deliverables are to be produced in terms of a series of reports that describe the project plan, the work carried out during the iterative process of data preparation, modelling and evaluation including data formatting, consistency or other quality issues, opportunities for useless instance or attribute removal and the approaches taken to solving issues with instances affected by noise, outliers or Missing values.

Stage 7 Tasks and Deliverables

Deployment
Take the evaluation results and determine a strategy for deployment. If a general procedure has been identified to create the relevant model(s), documented it here for later deployment.

Plan deployment
Evaluate the results and conclude with a strategy for deployment of the data mining result(s) into the business.
Summarize the deployment strategy, including necessary steps and how to perform them.
i.e

- Summarize deployable results
- Develop and evaluate alternative plans for deployment
- Decide for each distinct knowledge or information result
- Determine how knowledge or information will be propagated to users
- Decide how the use of the result will be monitored and its benefits measured (where applicable)

- Decide for each deployable model result
- Establish how the model result will be deployed within the organization's systems
- Determine how its use will be monitored and its benefits measured

Plan monitoring and maintenance

Monitoring and maintenance are important issues if the data mining results become part of the day-to-day business and its environment. A careful preparation of a maintenance strategy helps to avoid unnecessarily long periods of incorrect usage of data mining results.

Produce an **Monitoring and maintenance plan that** summarizes monitoring and maintenance strategy, including necessary steps and how to perform them.

- Check for dynamic aspects (i.e., what things could change in the environment?)
- Decide how accuracy will be monitored
- Determine when the data mining result or model should not be used any more. Identify criteria (validity,
- threshold of accuracy, new data, change in the application domain, etc.), and what should happen if the model or result could no longer be used. (update model, set up new data mining project, etc.).
- Will the business objectives of the use of the model change over time? Fully document the initial problem the model was attempting to solve.

Conclusion

Produce either a summary of the project and its experience, or a final presentation of the data Mining result(s).

Mark Allocation

Content	Weight
Plan deployment	30
Plan monitoring and maintenance	30
Conclusion	10
Total	**70**

Additional Information

- All work must be done on your own.
- Belgium Campus have software that can **scan for plagiarism** and a student caught doing this will get 0 for this assignment.
- Late assignments will not be accepted; missing the deadline is an automatic 0.