CRISP-DM on Job Readiness - Case study

Introduction:

This case study is based on Irish labour market.The outcome of this model is to determine whether a person is ready for employment or needs more training.CRISP-DM method was adopted in this case to create the model.There a 5 stages involved in this CRISP-DM..

1. Business Understanding:

In this stage,the current method is a case officer looks into the hundreds of applications and determines whether a client is ready to work or needs more training.To develop a model, customer needs are to be exactly matched.The goal of this case study is to analyse the unemployed customers,relationship between features and determine job readiness.

2. Data Preparation:

The real world dataset was used excluding sensitive information.Initial analysis were carried out on RapidMiner and SQL server.After discussions with the domain experts, 70 attributes were selected for further analysis.In data cleaning process only 8 from 70 attributes were clean while other columns were imputed with zeros,multiple zeros,none,no,NULL etc...Also the date attribute was stored as integer.Also we have do some processing to extract useful information from raw input.Also in interview data column most were empty so if we try to remove there were too much loss of data.Also a new column age was calculated from DOB.

3. Data Understanding:

Two methods were used to understand data Exploratory Statistics using Rapid Miner and Data Visualization using Tableau.This step is used to understand the hidden insights from the data by analysing each column.The ratio of men and women customers were almost same.The average age is 35.5 with minimum age is 16 and the maximum age is 101.Only 10% of the customers are dependent adult .Maximum number of children for a customer is 12 with average less than 1.Also majority of the clients were poorly educated with no skills,

Trends can be easily figured out using visualization.It is found that maximum number of customers are young and were registered in the year 2013 .Most of the clients,require prior work experience from the customer.The maximum number of registrations was in december and minimum was in the month of january.

4. Modelling :

The objective of this project is to apply various techniques to achieve consistency to determine the job readiness.It was decided to be a binomial classification problem using algorithms that provide most informative feedback.All processing were executed in RapidMiner.More variable were also analysed using correlation matrix.

Decision Tree:

The accuracy of decision tree model was **81.97%**.Experience(**EXPI**) was used to split.Again a decision tree model was created excluding the **EXPI** factor.Now experience declare by client irelating to secondary profession(**MANCO2**) and willingness for full time(**FULL_TIME)** job was figured out to be important variables.

Random Forest:

The accuracy of using 20 trees was **57.61%** and split parameters was same as decision tree

Support Vector Machine:

The accuracy while using **SVM** was **58.64%.**

K-Nearest Neighbour:

The accuracy while using **KNN** was **75.15%** and this also proved underlying pattern of experience with the result. Correlation matrix was used to analyse the findings of the classification process.Various weighting techniques was used to identify the attributes crucially affecting the job readiness.

5. Evaluation and Conclusion:

This method confirmed the existence of pattern among the unemployed population registered with the Irish employment services.There is straightforward relationship among the features of the customers,a person is job ready if they have work experience and willing to be employed full time.Also some cases ever person with work experience and willing to work full time was suggested some training.The conclusion from the case study was building a model is not going to take any advantage over the current method.Also more education programs and training are to be given to increase job readiness.

https://docs.google.com/document/d/1N22NHxOd3mtqNCcHupuBKFL1lurbvjjTrpRMJCXmezM/edit?usp=sharing