Marwen : slide 1-2-3-4

Aziz Slide 5-6

AHMED: Slide 7-8-9

Amine: Slide 10-11-12-13

Yahya Slide 14-15-16 -17

Noura Slide 18-19-20

Marwen conclusion

Intro

--------------------------------------------------------------------------------------------------------------------------------------

Slide 3

Auto insurance is a policy purchased by vehicle owners to mitigate costs associated with getting into an auto accident. Instead of paying out-of-pocket for auto accidents, people pay annual [premiums](https://www.investopedia.com/terms/p/premium.asp) to an auto [insurance](https://www.investopedia.com/terms/i/insurance.asp) company; the company then pays all or most of the costs associated with an auto accident or other vehicle damage.

Slide 4

There are 3 main goals any insurance company aims to achieve for their subscribers:

-protection:the client’s health and vehicule is safeguarded from harm through various policy types, such as collision insurance , Personal injury protection insurance , Underinsured motorist insurance etc…

-compensation :material damage and/or physical injuries are reimbursed monetarily , after the claim is processed by the insurance company

-justice: establishing a fair investigation process by trained agent to ensure that every claim is treated with absolute objectivity

Slide 5

Due to the exponential growth of the data and info accumulated that must processed by an insurance company, the CGA has encountered certain problems concerning data analysis in the automobile insurance domain, and some of these issues are:

-Certain calculations are done manually, which takes time and might give incorrect results

-the two previous problems lead to more cases of fraud, this point will be explained thoroughly in a later part.

Slide 6

Profit maximization:From the insurer’s point of view , the goal of an insurance company is to maximise profit.

Bonus Malus predection : one of the important aspect is the bonus-malus index (BM) which is given to a client when they start an insurance policy , most of the company’s financial loss comes from attributing the default Index to a client that is prone to making accidents

Fraud detection :another cause of financial loss are fraudulent claims , which take considerable time to be reviewed by agents and some can go unchecked

Slide 7

The primary database was provided by the CGA itself , it contains the tables shown on the right , this data is supplemented with the excel files shown to the left

The previously mentioned data will be used to assist in creating the predictive models,from now on it will be referenced as “internal data”, any other information outside this set is called “ external data”.

Slide 8

External data shown here was gathered from surveys sites concerning insurance companies satisfaction , this information can guide the CGA to focus on certain companies from a partnership point of view

Slide 9

Earlier this year , we launched our own survey which aims to to aid in recognizing churn reasons , among other intel , the pie chart shown here shows most common reasons

Slide 10

Before creating the predictive models , the information must first be prepared , to streamline this task , we created a data warehouse consisting of two main datamarts , one for predicting Nonus malus index , the second for detecting fraudulent activities.

Slide 11

The raw internal data shown in slide 7 was merged using SQL server integration services (SSIS).

The resulting tables was saved back in ssms to be accesed by the python scripts

Slide 12

Columns containing null data were deleted  
missing values are filled with proportionate data from the column

Certain variable types were changed to accommodate machine learning models

Slide 13

To ensure an accurate predictive output , a new column was added called “fraud”, it flags

Suspicious clieents and bonus malus indexes , these are disregarded from the data modelling phase

Slide 14

Once our internal data is cleaned and prepared it is time for the modelling phase. We noticed that our data is a labeled data with the target class being the Bonus Malus class thus we’re going to use supervised learning models. We have also noted that our data is discrete which is why we used several classification models and searched for the most adequate ones to our data such as KNN, Naïve Bayes SVM etc …

Slide 15

Decision Trees is a rule-based model, it used for classification and generally used for decision making and in forecasting future outcomes and assigning probabilities to those outcomes.

Slide 16

The evaluation of the model gave us an accuracy of 0.94

This model is chosen for its relative speed which assists in continuous update and also for its respectable accuracy

Slide 17

Slide 18

The final phase of this project is providing the user with a usable interface , it serves two goals , an interactive prediction form which correspond to our two main project goals , and a visual statstics dashboard

Slide 19

the prediction form is easy to read , easy to use interface that dosent necesstitates intensive technical know how ,

slide 20

this screenshot is a glimpse of the full statistics page , it provides various models such as geographical maps for car accident locations and dynamic line chart for number of accidents per state.

To provide comfort to the end user , we added some quality of life features such as dark/light mode.

Slide 21