一、Topic: Used Cars Data Exploratory Analysis and Prediction in Belarus

Research contents:

1. Get the datas
   (1) Resourse: https://www.kaggle.com/slavapasedko/belarus-used-cars-prices
   (2) data cleaning and preprocessing
      ① consistency check
      ② Analysis and processing of missing datas and outliers
      ③ devide datas into category datas and digital datas
2. Fundamental features description of the data
   (1) the distribution of the datas
      ① distribution diagrams
      ② skewness
      ③ kurtosis
   (2) use charts to visualise the datas
      ① histograms
      ② pie charts
      ③ bar charts
      ④ ...
3. Regression
   (1) Goal: used cars price prediction
   (2) process:
      ① correlation analysis
         a. heatmaps
         b. scatter plot
      ② according to correlation and t value, choose the independent variable, then do regression
      ③ model performance test
         a. R square
         b. mean absolute error
         c. mean Squared error
4. Classification
   (1) Goal: divide prices into three categories(cheap, average and expensive) by using 'car firm', 'conditon', 'mileage', etc.
   (2) Method:
      ① Random Forest
      ② K-NN
      ③ SVM
5. Evaluation
   (1) We divide the whole datas into two parts. We use 85% of the datas as train datas, and 15% of the datas for testing the regression accuracy and classification.

   (2) Method:
      ① Confusion Matrix

② ROC curve
③ Logarithmic Loss
④ Classification Accuracy

6. Conclusion

We use the outcomes gained from the models to make a conclusion, giving our perspectives about the used cars market in Belarus.

二、Personal Assignment:

1. Taohe Zhan - Fundamental features description of the data
2. Jie Huang - Regression
3. Nianqing Chen - Classificaition
4. Liuyi Pan - Evaluation