ALKAN ULUTAŞ

Project

1737285

Contents

[1. Homework Details 2](#_Toc498180431)

[2. Dataset 3](#_Toc498180432)

[2.1. Input Information 3](#_Toc498180433)

[2.2. Output Information 4](#_Toc498180434)

[3. Result 6](#_Toc498180435)

[4. Source Code 7](#_Toc498180436)

# Introduction

Study is about predicting the net profit/loss of companies by using financial statement of each them which is based on quarter term of years between 2013 and 2017.

It depends on implementing Linear Regression, Random Forest, Decision Table, Simple Linear Regression, M5P on dataset.

The aim of the dataset is to predict the net profit/loss of companies from financial statement of companies and next term will be more profitable or not. Originally it is a regression problem in which the output is the net profit/loss amount in years.

# Input Information

On this study, financial statement dataset is used which is got from [www.foreks.com](http://www.foreks.com) whose provided for study by serving webservice. Dataset has 923 rows and 80 attributes.

|  |  |  |
| --- | --- | --- |
| feature 1 | Working Assets | numeric |
| feature 2 | Cash and Cash Equivalents | numeric |
| feature 3 | Trade Receivables | numeric |
| feature 4 | trade receivables from affiliates | numeric |
| feature 5 | trade receivables from nonaffiliated | numeric |
| feature 6 | receivables from financial activities | numeric |
| feature 7 | receivables from non-related parties of financial sector operations | numeric |
| feature 8 | other receivables | numeric |
| feature 9 | other receivables from related parties | numeric |
| feature 10 | other receivables from unrelated parties | numeric |
| feature 11 | derivative financial instruments | numeric |
| feature 12 | stocks | numeric |
| feature 13 | biological asset | numeric |
| feature 14 | prepaid expenses | numeric |
| feature 15 | assets related to current period tax | numeric |
| feature 16 | other current assets | numeric |
| feature 17 | subtotal | numeric |
| feature 18 | assets classified as held for sale | numeric |
| feature 19 | real assets | numeric |
| feature 20 | financial investments | numeric |
| feature 21 | other receivables | numeric |
| feature 22 | other receivables from related parties | numeric |
| feature 23 | other receivables from unrelated parties | numeric |
| feature 24 | investments valued by equity method | numeric |
| feature 25 | tangible assets | numeric |
| feature 26 | intangible assets | numeric |
| feature 27 | special assement | numeric |
| feature 28 | other intangible assets | numeric |
| feature 29 | prepaid expenses | numeric |
| feature 30 | deferred tax assets | numeric |
| feature 31 | other fixed assets | numeric |
| feature 32 | total assets | numeric |
| feature 33 | resources | numeric |
| feature 34 | short-term liabilities | numeric |
| feature 35 | short-dated loan | numeric |
| feature 36 | short-term portions of long-term borrowings | numeric |
| feature 37 | other financial liabilities | numeric |
| feature 38 | trade payables | numeric |
| feature 39 | trade payables to related parties | numeric |
| feature 40 | trade payables to unrelated parties | numeric |
| feature 41 | liabilities under employee benefits | numeric |
| feature 42 | other liabilities | numeric |
| feature 43 | other liabilities to related parties | numeric |
| feature 44 | other liabilities to unrelated parties | numeric |
| feature 45 | derivative financial instruments | numeric |
| feature 46 | government incentives and grants | numeric |
| feature 47 | deferred incomes | numeric |
| feature 48 | current income tax liability | numeric |
| feature 49 | short-term provisions | numeric |
| feature 50 | short-term provisions related to employee benefits | numeric |
| feature 51 | other short-term provisions | numeric |
| feature 52 | other short term liabilities | numeric |
| feature 53 | subtotal | numeric |
| feature 54 | liabilities of asset groups classified as sales | numeric |
| feature 55 | long-term liabilities | numeric |
| feature 56 | long-term borrowings | numeric |
| feature 57 | long-term provisions | numeric |
| feature 58 | long-term provisions for employee benefits | numeric |
| feature 59 | other long-term provisions | numeric |
| feature 60 | liabilities related to current period tax | numeric |
| feature 61 | deferred tax liabilities | numeric |
| feature 62 | other long-term liabilities | numeric |
| feature 63 | owner’s equity | numeric |
| feature 64 | equity of the parent company | numeric |
| feature 65 | paid capital | numeric |
| feature 66 | capital adjustment differences | numeric |
| feature 67 | recovered shares | numeric |
| feature 68 | premiums and discounts on shares | numeric |
| feature 69 | other comprehensive income and expenses that will not be reclassified to profit or loss | numeric |
| feature 70 | revaluation and measurement gain / losses | numeric |
| feature 71 | other earnings / losses | numeric |
| feature 72 | other comprehensive income and expenses that will be reclassified to profit or loss | numeric |
| feature 73 | foreign currency translation differences | numeric |
| feature 74 | hedonic gain / losses | numeric |
| feature 75 | revaluation and classification gains / losses | numeric |
| feature 76 | other earnings / losses | numeric |
| feature 77 | restricted reserves | numeric |
| feature 78 | profits / losses of previous years | numeric |
| feature 79 | net period profit / loss | numeric |
| feature 80 | total resources | numeric |

# Output Information

Total Number of Instances: 923

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Linear Reg. | Simple Linear Reg. | Random Forest | Decision Table |  |
| Correlation coefficient | 0.6296 | 0.9319 | 0.8909 | 0.8031 |  |
| Mean absolute error | 222367692091.5847 | 117196962604.2879 | 124157156523.3976 | 172294399183.0121 |  |
| Root mean squared error | 781839916386.1993 | 291690805639.2097 | 365814029529.5052 | 481031357384.9382 |  |
| Relative absolute error | 68.3153 % | 36.005 % | 38.1433 % | 52.9319 % |  |
| Root relative squared error | 97.14 % | 36.2412 % | 45.4507 % | 59.7659 % |  |

# Result

It is not possible to say that when i increase the count of train data rows, accuracy of our model doesn’t increase together.

Using all features by 2000 raws of training data has highest accuracy at the end of the my all calculations.

# References

Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt 35, Sayı 1, 2017, s. 67-88, Mehmet ÖZÇALICI

Hisse Senedi Fiyat Tahminlerinde Bilgi İşlemsel Zeka Yöntemleri, Mehmet ÖZÇALICI

Stock Prediction using Machine Learning a Review Paper, Nirbhey Singh Pahwa

Machine Learning Techniques for Stock Prediction, Vatsal H. Shah