

**UNIVERSITY OF ECONOMICS AND LAW**

**VNUHCM**

**END-OF-COURSE PROJECT:**

**PROGRAM PACKAGE IN FINANCE 2**

**YEAR 2023**

**<DEBT MATURITY STRUCTURE OF THANG LONG INVESTMENT GROUP JOINT STOCK COMPANY>**

**Syllabus: K20414C\_ Fintech**

**Course: 222CN0901**

**Student: Lương Thị Mỹ Tâm**

**MSSV: K204141929**

**UNIVERSITY OF ECONOMICS AND LAW**

**VNUHCM**

**END-OF-COURSE PROJECT:**

**PROGRAM PACKAGE IN FINANCE 2**

**YEAR 2023**

**<DEBT MATURITY STRUCTURE OF THANG LONG INVESTMENT GROUP JOINT STOCK COMPANY>**

**DECLARATION**

I hereby declare that the end-of-course project "Debt maturity structure of Thang Long investment group JSC" is the result of my work under the guidance of Dr. Nguyen Thanh Liem, within the framework of the subject 'Program package in finance 2'.

**TABLE OF CONTENTS**

[DECLARATION 1](#_Toc7542)

[TABLE OF CONTENTS 2](#_Toc16515)

[LIST OF TABLES/FIGURES 4](#_Toc24223)

[TASK 1: Perform literature review 7](#_Toc26996)

[TASK 2: Data collection and input 9](#_Toc26791)

[TASK 3: Provide descriptive statistics of all the variables 9](#_Toc24170)

[A) Entire period 9](#_Toc27585)

[B) Before period 10](#_Toc24010)

[C) After period 10](#_Toc18356)

[TASK 4: Provide box & whisker plot and histogram of debt maturity structure 12](#_Toc24404)

[1. Entire Period 12](#_Toc6160)

[2. Before Period 13](#_Toc12281)

[3. After Period 14](#_Toc7956)

[TASK 5: Perform multiple regression to determine the significant determinants of debt maturity structure 15](#_Toc30127)

[A) Model 1 15](#_Toc27855)

[B) Model 2 17](#_Toc8839)

[C) Using model 1 to predict debt maturity structure 19](#_Toc19524)

[TASK 6: Using Arima to predict debt maturity structure in 2022 20](#_Toc22993)

[A) Checking stationary and fixing by taking the difference 20](#_Toc3725)

[B) Finding optimal parameter 21](#_Toc7963)

[C) Predicting 21](#_Toc19551)

[D) Evaluation matrix 22](#_Toc9789)

[E) Check white noise 22](#_Toc14421)

[F) Impoving arima model 23](#_Toc5345)

[Task 7. Explain in fewer than 150 words how Decision Tree algorithm can be used to make prediction whether the firm will increase/decrease the debt maturity structure 24](#_Toc18722)

[REFERENCE 24](#_Toc5680)

[APPENDIX 25](#_Toc2058)

# **LIST OF TABLES/FIGURES**

|  |  |  |
| --- | --- | --- |
|  | **Name of tables/figures** | **Page** |
| 1 | Table 1: The Data set | 9 |
| 2 | Table 2:  Descriptive Statistics of entire period | 9 |
| 3 | Table 3:  Descriptive Statistics of before period | 10 |
| 4 | Table 4:  Descriptive Statistics of  after period | 10 |
| 5 | Figure 5: The box plot of entire period | 12 |
| 6 | Figure 6: The histogram chart of entire period | 13 |
| 7 | Figure 7: The box plot of before period | 13 |
| 8 | Figure 8: The histogram chart of before period | 14 |
| 9 | Figure 9: The box plot of after period | 14 |
| 10 | Figure 10: The histogram chart of after period | 15 |
| 11 | Table 11: The result of regression model 1 | 16 |
| 12 | Table 12: Checking multicollinearity and heteroskedasticity | 17 |
| 13 | Table 13: The result of regression model 2 | 18 |
| 14 | Table 14: Checking multicollinearity and heteroskedasticity | 19 |
| 15 | Figure 15: The chart of forecast results vs reality | 20 |
| 16 | Table 16: The stationarity test results | 21 |
| 17 | Table 17: The optimal model results | 21 |
| 18 | Figure 18: The chart of forecast results vs reality | 22 |
| 19 | Figure 19: The evaluation matrix | 22 |
| 20 | Figure 20: The result of white noise test | 23 |
| 21 | Figure 21: The chart of forecast results vs reality | 23 |

**TASK 1: Perform literature review**

Terra (2011) conducted a study with a data set of 1693 non-financial firms from the seven largest economies of Latin America and companies from the US over 16-years period to discover factors affecting debt maturity structure. The study used Dynamic panel data analysis estimated by the generalized method of moments. The results indicate that Liquidity and Lagged Maturity have a positive effect on maturity structure, while Tax effects has a negative effect. Leverage has only a positive effect on US companies while ROA, Growth opportunities, and Tangibility are not statistically significant.

Pham Thi Van Trinh (2018) examined the debt maturity structure of 67 real estate companies listed on the Vietnamese stock market in the period from 2010 to 2016. The study used the Sys-GMM model, the research results show that financial leverage, firm size, asset structure, liquidity, and profit volatility are the factors that positively affect the debt maturity structure of enterprises, except profitability is the opposite effect. Other factors such as growth opportunities, and corporate income tax are not statistically significant.

Kalsie & Nagpal's study (2018) investigated the firm-specific and macroeconomic determinants that had an impact on the debt maturity structure of Indian corporate firms including 29 non-financial firms listed on the National Stock Exchange in the period from 2008 to 2016. The study used panel data analysis. The results indicate that firm size, liquidity, asset maturity, and prime interest rates have a significant influence positively on the choice of debt maturities in India, while the effects of taxes, growth rates, firm quality, and wholesale price index are not significantly related to debt maturity structure.

Mohammed (2020) analyzed data taken from annual reports covering 92 listed non-financial firms in Nigeria between 2010 and 2015 using the Two-stage Generalised Method of Moments (GMM) regression model. The results show that diversification, growth opportunity, liquidity, and asset intensity are significant negative affect. Dividend policy, firm size, and non-debt tax shield were found to be significant positive determinants of debt maturity structure among the listed firms in Nigeria. However, this study did not document significant evidence in favor of profitability (ROA) and investors’ confidence as determinants of debt maturity structure even though they are among the determinants of debt maturity choices in some other studies with strong theoretical backing.

Thi Van Trang Do (2021) used data collected from consumer goods companies listed on Vietnam's stock exchange from 2009 to 2019 to analyze factors affecting debt maturity structure. The paper uses the feasible generalized least squares (FGLS) estimation. The empirical results indicate that microeconomic factors, such as capital structure, asset structure, asset liquidity, and firm size, have positively influenced debt maturity structure and are statistically significant while profitability has a negative effect. Meanwhile, macroeconomic factors such as inflation rate and credit growth have a significant effect on corporate debt maturity.

Kim Quoc Trung Nguyen (2022) studied the factors affecting the debt maturity structure of 176 non-financial small and medium-sized companies listed in Vietnam in the period from 2010 to 2019. The results show that debt maturity structure is positively correlated with the lagged debt maturity structure, firm size, growth opportunities and GDP under agency cost theory, while signaling theory creates the ground for the negative effects of the profitability and leverage on debt maturity structure under the GMM method.

Based on literature review above, I decided to choose four independent variables are expected to positively affect Debt Maturity Structure for the model as follows:

+ Firm Size: According to agency theory, small firms tend to have higher agency costs. These costs can be managed through the use of short-term financing. This implies that there is a positive relationship between firm size and the maturity structure of debt. It calculated as ln (total assets)

+ Liquidity: The signaling theory suggests that businesses with high liquidity should use more long-term debt to minimize the cost of capital. It calculated as the ratio of current assets to current liabilities

+ Leverage: The signaling theory suggests that businesses with high levels of debt should use long-term debt to reduce liquidity risk and minimize the risk of bankruptcy. It calculated as the ratio of total liabilities to total assets.

+ ROA: represent the company's profitability. The signaling theory suggests that there is a negative correlation between the maturity structure of debt and profitability.

**TASK 2: Data collection and input**

|  |
| --- |
| K204141929 <- read\_excel("K204141929.xlsx")  View(K204141929)  DMS <- K204141929[3]/K204141929[4]  Liquidity <- K204141929[5]/K204141929[6]  Size <- log(K204141929[7])  Leverage <- K204141929[8]/K204141929[7]  ROA <- K204141929[9]  dt <- data.frame(K204141929[1], K204141929[2],DMS, Liquidity, Size, Leverage, ROA)  dt <- setNames(dt, c("Code","Date","DMS","Liquidity","Size","Leverage","ROA"))  options(scipen=999)  View(dt) |

|  |
| --- |
|  |
| Table 1: The Data set |

**TASK 3: Provide descriptive statistics of all the variables**

1. **Entire period**

|  |
| --- |
| sta <- data.frame(min=sapply(dt[3:7], min, na.rm = TRUE),  max=sapply(dt[3:7], max, na.rm = TRUE),  mean=sapply(dt[3:7], mean, na.rm = TRUE),  median= sapply(dt[3:7], median, na.rm = TRUE),  sd=sapply(dt[3:7], sd, na.rm = TRUE))  View(sta) |

|  |
| --- |
|  |

Table 2: Descriptive Statistics of entire period

1. **Before period**

|  |
| --- |
| dt1 <- data.frame(select(dt, c(1,2,3,4,5,6,7))  %>% filter(dt[2] < "2020-03-31"))  dt1 <- setNames(dt1, c("Code","Date","DMS","Liquidity","Size","Leverage","ROA"))  View(dt1)  sta1 <- data.frame(min=sapply(dt1[3:7], min, na.rm = TRUE),  max=sapply(dt1[3:7], max, na.rm = TRUE),  mean=sapply(dt1[3:7], mean, na.rm = TRUE),  median=sapply(dt1[3:7], median, na.rm = TRUE),  sd=sapply(dt1[3:7], sd, na.rm = TRUE))  View(sta1) |

|  |
| --- |
|  |

Table 3: Descriptive Statistics of before period

1. **After period**

|  |
| --- |
| dt2 <- data.frame(select(dt, c(1,2,3,4,5,6,7))  %>% filter(dt[2] >= "2020-03-31"))  dt2 <- setNames(dt2, c("Code","Date","DMS","Liquidity","Size","Leverage","ROA"))  View(dt2)  sta2 <- data.frame(min=sapply(dt2[3:7], min, na.rm = TRUE),  max=sapply(dt2[3:7], max, na.rm = TRUE),  mean=sapply(dt2[3:7], mean, na.rm = TRUE),  median=sapply(dt2[3:7], median, na.rm = TRUE),  sd=sapply(dt2[3:7], sd, na.rm = TRUE))  View(sta2) |

|  |
| --- |
|  |

Table 4: Descriptive Statistics of after period

**Comment:**

* The descriptive statistics results shown in the 3 tables above indicate that during the entire period, on average, the company used about 19.27% long-term debt in total debt with a median value of 0.1083 and the standard deviation shows that the variability of long-term debt compared to the average is 22.49%, which is higher than the average value. In addition, the minimum value of the debt maturity structure is 0.00009, which is very small and belongs to the after period, indicating that at some point the company used very little long-term debt while the company used the most long-term debt at 74.31% belonging to the before period. And when comparing all 3 periods, we see that the before period is the period when the company has the largest average debt maturity structure (0.2292) and the variability in this period is also the strongest (0.2345) while the after period is the period when the company uses the least long-term debt (0.0332) and has little variability (0.0235).
* For Liquidity, the highest average value is 3.6111 belonging to the before period, however, the entire period is the period with the strongest variability (1.3221). The company has the lowest liquidity value of 1.5423 belonging to the after period and the highest value of 5.9348 belonging to the before period.
* For Firm size, the average size of the company in the entire period is at 13.53, where the size in the before period is 13.03 smaller than in the after period which is 14.51. However, the variability of size in the entire period is largest at 0.747. The largest value of size is 15.07 belonging to the after period and the smallest value is 12.43 belonging to the before period.
* For Leverage, the average value of the company’s financial leverage in the entire period is 0.3426, meaning that on average about 34.26% of the company’s current assets are financed by current liabilities, with the highest use being 55.81% belonging to the after period and the lowest being 16.81% in the before period. The after period is the period when the company uses the highest average financial leverage of 43.47% and the before period is the lowest with 32.15%. The standard deviation of the entire period is the largest when compared to before and after periods.
* For ROA representing the company’s profitability, we see that the company’s lowest profitability is -0.0397 and highest is 0.1558 both in the before period. The average value in the entire period is 0.052, whereas the after period has an average ROA value of 0.067 higher than the before period which is 0.048. The standard deviation of ROA in the before period is highest at 0.046, and lowest at 0.009 in the after period.

=> From this, we can see the significant impact of Covid-19 on the company. It has caused the company’s debt maturity structure to decrease significantly and become less volatile, and liquidity has also decreased. However, the company’s size, financial leverage, and ROA have increased. This shows that the company has shifted to using mostly short-term debt, reducing long-term debt so its liquidity will be more risky, the company may face liquidity problems but this helps the company expand its size and increase its short-term profitability.

**TASK 4: Provide box & whisker plot and histogram of debt maturity structure**

1. **Entire Period**

|  |
| --- |
| dt %>%  filter(!is.na(DMS)) %>%  ggplot(aes(y = DMS,fill='DMS')) +  geom\_boxplot() +  coord\_flip()+  scale\_fill\_manual(values = c('#265493'))+  scale\_y\_continuous(labels = scales::comma) |

|  |
| --- |
|  |
| Figure 5: The box plot of entire period |

|  |
| --- |
| ggplot(dt, aes(x = DMS, fill='DMS')) +  geom\_histogram()+  scale\_fill\_manual(values = c('#265493'))+  geom\_histogram(binwidth = 0.05) +  scale\_x\_continuous(labels = scales::comma) |

|  |
| --- |
|  |

Figure 6: The histogram chart of entire period

1. **Before Period**

|  |
| --- |
| dt1 %>%  filter(!is.na(DMS)) %>%  ggplot(aes(y = DMS,fill='DMS')) +  geom\_boxplot() +  coord\_flip()+  scale\_y\_continuous(labels = scales::comma) |

|  |
| --- |
|  |

Figure 7: The box plot of before period

|  |
| --- |
| ggplot(dt1, aes(x = DMS, fill='DMS')) +  geom\_histogram()+  geom\_histogram(binwidth = 0.05) +  scale\_x\_continuous(labels = scales::comma) |

|  |
| --- |
|  |

Figure 8: The histogram chart of before period

1. **After Period**

|  |
| --- |
| dt2 %>%  filter(!is.na(DMS)) %>%  ggplot(aes(y = DMS,fill='DMS')) +  geom\_boxplot() +  coord\_flip()+  scale\_fill\_manual(values = c('yellow'))+  scale\_y\_continuous(labels = scales::comma) |

|  |
| --- |
|  |
| Figure 9: The box plot of after period |

|  |
| --- |
| ggplot(dt2, aes(x = DMS, fill='DMS')) +  geom\_histogram()+  scale\_fill\_manual(values = c('yellow'))+  geom\_histogram(binwidth = 0.05) +  scale\_x\_continuous(labels = scales::comma) |

|  |
| --- |
|  |

Figure 10: The histogram chart of after period

**Comment:**

* For the box plot:

We can see that all 3 periods have outliers, with the entire period having the most outliers and the before period having the fewest outliers. The median value of the before period is the highest while the after period is the lowest, consistent with the descriptive statistics results. Looking at the length of the box body, the before period has the longest length, indicating the greatest variability and the after period has the least variability. In addition, looking at the extreme values at the ends of the whiskers, we see that the before period has the largest range of points compared to the other two periods, indicating that the before period has the widest distribution and the after period has the narrowest distribution. In terms of skewness, all 3 periods are right-skewed.

* For the histogram:

All three periods have a right-skewed distribution with the presence of outliers. The company’s debt maturity structure is mainly concentrated in the range from 0 to 0.3, indicating that the company uses more short-term debt than long-term debt. This is especially evident in the after period when the company’s debt maturity structure fluctuates only in the range from 0.15 downwards. However, there are also many quarters where the company uses more than 50% long-term debt.

**TASK 5: Perform multiple regression to determine the significant determinants of debt maturity structure**

1. **Model 1**

* **Multiple regression**

|  |
| --- |
| summary(model1<-lm(DMS ~ Liquidity + Leverage + Size + ROA , data = dt)) |

|  |
| --- |
|  |

Table 11: The result of regression model 1

**Comment:**

Based on the results of the linear regression model, we see that the adjusted R-squared is 0.4568, meaning that 45.68% of the variation in the company’s debt maturity structure is explained by the independent variables. This suggests that this model has a relatively good predictive ability. All variables are statistically significant with p-values less than 0.1. In task 1, according to Terra (2011) and Mohammed (2020), the ROA variable is not statistically significant, however, the regression results show that the company’s ROA is statistically significant at a significance level of 10%. The two variables Liquidity and Leverage have a positive relationship with debt maturity structure while Size and ROA have a negative impact. These relationships are as expected by the author and support the research results of Pham Thi Van Trinh (2018), Kalsie & Nagpal’s study (2018), and Thi Van Trang Do (2021), except for Size which goes against expectations and results of the studies reviewed in task 1. In terms of coefficient significance, we can explain that when liquidity increases by 1 unit, its debt maturity structure will increase on average by 0.06787 units, other factors remaining constant (similarly for Leverage). For Size, we can explain that when size increases by 1 unit, the average value of DMS is expected to decrease by 0.07940 units, other factors remaining constant (similarly for ROA). The F-statistic is 9.831 with a p-value of 0.00001465, indicating that the model is statistically significant at a significance level of 0.05. This means that there is sufficient evidence to assert that at least one of the independent variables affects the dependent variable.

* **Test of multicollinearity and heteroskedasticity**

|  |
| --- |
| car::vif(model1)  bptest(model1)  shapiro.test(resid(model1)) |

|  |
| --- |
| #multicollinearity    #heteroskedasticity |

Table 12: Checking multicollinearity and heteroskedasticity

**Comment:**

* The VIF coefficients of all four variables are less than 10, so the model does not have multicollinearity.
* The p-value of Breusch-Pagan test is greater than 0.05 so the model is not heteroskedasticity. Besides, the p-value of Shapiro-wilk normality test is greater than 0.05 so we conclude that the data set is normally distributed.

1. **Model 2**

* **Multiple regression**

|  |
| --- |
| dt$Covid <- ifelse(dt$Date < "2020-03-31", 0, 1)  summary(model2 <-lm(DMS ~ Liquidity + Leverage + Size + ROA  + Covid\*Liquidity + Covid\*Size + Covid\*Leverage + Covid\*ROA , data = dt)) |

|  |
| --- |
|  |

Table 13: The result of regression model 2

**Comment:**

Based on the results table above, we see that the adjusted R-squared is 0.5045, meaning that 50.45% of the variation in the company’s debt maturity structure is explained by the independent variables. In the model, only the Liquidity, Leverage, and ROA variables are statistically significant. The Liquidity, Leverage, Size, Covid and ROA\*Covid interaction variables have a positive relationship with debt maturity structure while the ROA variable and the remaining interactive variables have a negative impact. In terms of coefficient significance, we can explain that when a company’s liquidity increases by 1 unit, its debt maturity structure will increase on average by 0.07494 units, with other factors unchanged (similar for Leverage, Size). For ROA we can explain that when ROA increases by 1 unit, the average expected value of DMS will decrease by 2.06442 units, with other factors unchanged. The coefficient for the Covid \* Liquidity interactive variable is -0.08284. This means that when there is Covid, the slope of the relationship between DMS and Liquidity will decrease by 0.08284 units (similar for Covid \*Size and Covid \*Leverage). And the coefficient for the Covid: ROA interactive variable is 1.73926. This means that when there is Covid, the slope of the relationship between DMS and ROA will increase by 1.73926 units. However, none of the interactive variables are statistically significant, which means that there is no evidence of an interaction effect between Covid and any of the other independent variables. The F-statistic is 5.751 with a p-value of 0.00008616 which is very small. This means that there is sufficient evidence to assert that at least one of the independent variables affects the dependent variable.

* **Test of multicollinearity and heteroskedasticity**

|  |
| --- |
| car::vif(model2)  bptest(model2)  shapiro.test(resid(model2)) |

|  |
| --- |
| #multicollinearity  #heteroskedasticity |

Table 14: Checking multicollinearity and heteroskedasticity

**Comment:**

* The VIF coefficients of the Covid variable and the interaction variables are all greater than 10, so the model exhibits multicollinearity..
* The p-value of Breusch-Pagan test is greater than 0.05 so the model is not heteroskedasticity. Besides, the p-value of Shapiro-wilk normality test is greater than 0.05 so we conclude that the data set is normally distributed.

1. **Using model 1 to predict debt maturity structure**

|  |
| --- |
| predictions <- predict(model1, newdata = dt)  predictions  dt$Predicted <- predictions  View(dt)  plot(as.vector(unlist(dt[3])), type = 'l', ylab = 'Value', col ="blue")  lines(as.vector(unlist(dt$Predicted)), lty = 'dotted', col='red') |

|  |
| --- |
|  |

Figure 15: The chart of forecast results vs reality

**TASK 6: Using Arima to predict debt maturity structure in 2022**

1. **Checking stationary and fixing by taking the difference**

|  |
| --- |
| DMSts = ts(dt$DMS, start =c(2010, 6), end =c(2019, 12), frequency = 4)  DMSts  #Check stationary  plot(DMSts, type = "l", main = "Debt Maturity Structure",  , xlab = "Year")  adf.test(DMSts)  pp.test(DMSts)  DMSts.diff <- diff(DMSts)  plot(DMSts.diff, type = 'l')  adf.test(DMSts.diff)  pp.test(DMSts.diff)  adf.test(diff(DMSts.diff))  pp.test(diff(DMSts.diff)) |

|  |
| --- |
|  |

Table 16: The stationarity test results

1. **Finding optimal parameter**

|  |
| --- |
| j <- forecast:::auto.arima(diff(DMSts.diff),ic='aic',trace=TRUE)  j |

|  |
| --- |
|  |

Table 17: The optimal model results

1. **Predicting**

|  |
| --- |
| forecasted <- forecast:::forecast.Arima(j, h=4,level=c(99.5))  forecasted  X2022 <- read\_excel("2022.xlsx")  plot(as.vector(unlist(X2022[2])), type = 'l', ylab = 'Value', col ="blue")  lines(as.vector(unlist(forecasted$mean[1:4])), lty = 'dotted', col='red') |

|  |
| --- |
|  |

Figure 18: The chart of forecast results vs reality

1. **Evaluation matrix**

|  |
| --- |
| X2022$forecast <- forecasted$mean[1:4]  y <- as.vector(unlist(X2022[2]))  x<- as.vector(unlist(X2022$forecast))  # Mean Absolute Error (MAE)  mae <- mean(abs(y - x))  # Mean Squared Error (MSE)  mse <- mean((y - x)^2)  # Root Mean Squared Error (RMSE)  rmse <- sqrt(mse)  # Mean Absolute Percentage Error (MAPE)  mape <- mean(abs((y - x) / y)) \* 100  mae  mse  rmse  mape |

|  |
| --- |
|  |

Figure 19: The evaluation matrix

1. **Check white noise**

|  |
| --- |
| Box.test(diff(DMSts.diff), lag = 2, type = "Ljung-Box")  shapiro.test(diff(DMSts.diff)) |

|  |
| --- |
|  |

Figure 20: The result of white noise test

1. **Impoving arima model**

|  |
| --- |
| t <- arima(DMSts, order =c(0,0,3))  t  k <- forecast:::forecast.Arima(t, h=4,level=c(99.5))  k  plot(as.vector(unlist(X2022[2])), type = 'l', ylab = 'Value', col ="blue")  lines(as.vector(unlist(k$mean[1:4])), lty = 'dotted', col='red') |

|  |
| --- |
|  |

Figure 21: The chart of forecast results vs reality

**Comment:**

First, the DMS variable will be converted to time-series data and then the stationary of the data will be checked. The result shows that the data is not stationary so we will fix it by taking the difference with the second difference (p-value <0.05) the data is stable. Next, we will find the optimal parameter with auto.arima and the result shows that the best model is (3,0,1). Then, fit the model and forecast 4 quarters in 2022. The forecast result compared to the actual data is shown in the chart above and evaluated through MAE, MSE, RMSE, and MAPE indices and when checking white noise there is no white noise (p-value < 0.05), however, the forecast result is not very good so we proceed to improve the model by changing the p, d, q parameters based on the ACF and PACF charts and the result shows that the (0,0,3) model is suitable and when using this model to forecast, the result is much better than before, closer to reality.

**Task 7. Explain in fewer than 150 words how Decision Tree algorithm can be used to make prediction whether the firm will increase/decrease the debt maturity structure**

The decision tree algorithm can be used to predict whether a firm will increase or decrease its debt maturity structure by analyzing the existing financial data of the firm. The algorithm works by building a tree-like model of decisions and their potential consequences based on historical data. The first step is to identify the relevant features that affect the firm's debt maturity structure. Once these features have been identified, the algorithm builds a model in which each node represents a decision based on one of the features. The model is trained using historical data to determine the optimal decision at each node. Finally, To make a prediction, the algorithm starts at the root of the tree and follows the path that matches the current state of the company. When it arrives at a leaf node, it provides a prediction based on the probabilities associated with that outcome. By inputting the relevant feature data into the model, the algorithm can predict whether the firm will increase or decrease its debt maturity structure. This information can be used by investors and other stakeholders to make informed decisions about the firm's financial health and prospects.

**REFERENCE**

Do, T. V. T. (2021). DETERMINANTS OF CORPORATE DEBT MATuRITY: EVIDENCE FROM THE CONSUMER GOODS SECTOR IN VIETNAM. *Innovations*, *18*(3), 175-182.

Kalsie, A., & Nagpal, A. (2018). The determinants of corporate debt maturity for nse-listed corporates. *FIIB Business Review*, *7*(1), 43-56.

Mohammed, L. (2020). Firm-specific determinants of debt maturity structure of listed non-financial firms in Nigeria. *Malaysian management journal*.

Nguyen, K. Q. T. (2022). Determinants of debt maturity structure: Evidence in Vietnam. *Cogent Business & Management*, *9*(1), 2094588.

Terra, P. R. S. (2011). Determinants of corporate debt maturity in Latin America. *European Business Review*, *23*(1), 45-70.

Vân Trinh, P. T. (2018). Cấu trúc kỳ hạn nợ của các công ty kinh doanh bất động sản niêm yết trên thị trường chứng khoán Việt Nam. *TẠP CHÍ KHOA HỌC ĐẠI HỌC MỞ THÀNH PHỐ HỒ CHÍ MINH-KINH TẾ VÀ QUẢN TRỊ KINH DOANH*, *13*(1), 38-50.

**APPENDIX**

install.packages('dplyr')

library(dplyr)

library(readxl)

#Task 2. Data collection and input

K204141929 <- read\_excel("K204141929.xlsx")

View(K204141929)

DMS <- K204141929[3]/K204141929[4]

Liquidity <- K204141929[5]/K204141929[6]

Size <- log(K204141929[7])

Leverage <- K204141929[8]/K204141929[7]

ROA <- K204141929[9]

dt <- data.frame(K204141929[1], K204141929[2],DMS, Liquidity, Size, Leverage, ROA)

dt <- setNames(dt, c("Code","Date","DMS","Liquidity","Size","Leverage","ROA"))

options(scipen=999)

View(dt)

summary(dt)

sum(is.na(dt))

#Task 3. Provide descriptive statistics of all the variables

#Entire period

sta <- data.frame(min=sapply(dt[3:7], min, na.rm = TRUE),

max=sapply(dt[3:7], max, na.rm = TRUE),

mean=sapply(dt[3:7], mean, na.rm = TRUE),

median= sapply(dt[3:7], median, na.rm = TRUE),

sd=sapply(dt[3:7], sd, na.rm = TRUE))

View(sta)

#Before period

dt1 <- data.frame(select(dt, c(1,2,3,4,5,6,7))

%>% filter(dt[2] < "2020-03-31"))

dt1 <- setNames(dt1, c("Code","Date","DMS","Liquidity","Size","Leverage","ROA"))

View(dt1)

sta1 <- data.frame(min=sapply(dt1[3:7], min, na.rm = TRUE),

max=sapply(dt1[3:7], max, na.rm = TRUE),

mean=sapply(dt1[3:7], mean, na.rm = TRUE),

median=sapply(dt1[3:7], median, na.rm = TRUE),

sd=sapply(dt1[3:7], sd, na.rm = TRUE))

View(sta1)

#After period

dt2 <- data.frame(select(dt, c(1,2,3,4,5,6,7))

%>% filter(dt[2] >= "2020-03-31"))

dt2 <- setNames(dt2, c("Code","Date","DMS","Liquidity","Size","Leverage","ROA"))

View(dt2)

sta2 <- data.frame(min=sapply(dt2[3:7], min, na.rm = TRUE),

max=sapply(dt2[3:7], max, na.rm = TRUE),

mean=sapply(dt2[3:7], mean, na.rm = TRUE),

median=sapply(dt2[3:7], median, na.rm = TRUE),

sd=sapply(dt2[3:7], sd, na.rm = TRUE))

View(sta2)

#Task 4.Provide box & whisker plot and histogram of the variable of assigned topic

#Entire period

library(ggplot2)

dt %>%

filter(!is.na(DMS)) %>%

ggplot(aes(y = DMS,fill='DMS')) +

geom\_boxplot() +

coord\_flip()+

scale\_fill\_manual(values = c('#265493'))+

scale\_y\_continuous(labels = scales::comma)

ggplot(dt, aes(x = DMS, fill='DMS')) +

geom\_histogram()+

scale\_fill\_manual(values = c('#265493'))+

geom\_histogram(binwidth = 0.05) +

scale\_x\_continuous(labels = scales::comma)

#Before period

dt1 %>%

filter(!is.na(DMS)) %>%

ggplot(aes(y = DMS,fill='DMS')) +

geom\_boxplot() +

coord\_flip()+

scale\_y\_continuous(labels = scales::comma)

ggplot(dt1, aes(x = DMS, fill='DMS')) +

geom\_histogram()+

geom\_histogram(binwidth = 0.05) +

scale\_x\_continuous(labels = scales::comma)

#After period

dt2 %>%

filter(!is.na(DMS)) %>%

ggplot(aes(y = DMS,fill='DMS')) +

geom\_boxplot() +

coord\_flip()+

scale\_fill\_manual(values = c('yellow'))+

scale\_y\_continuous(labels = scales::comma)

ggplot(dt2, aes(x = DMS, fill='DMS')) +

geom\_histogram()+

scale\_fill\_manual(values = c('yellow'))+

geom\_histogram(binwidth = 0.05) +

scale\_x\_continuous(labels = scales::comma)

#Task 5.Perform multiple regression to determine the significant determinants of the variable of assigned topic.

#Model 1

summary(model1<-lm(DMS ~ Liquidity + Leverage + Size + ROA, data = dt))

car::vif(model1)

par(mfrow=c(2,2))

plot(model1)

install.packages('lmtest')

library(lmtest)

bptest(model1)

shapiro.test(resid(model1))

#Model 2

dt$Covid <- ifelse(dt$Date < "2020-03-31", 0, 1)

summary(model2 <-lm(DMS ~ Liquidity + Leverage + Size + ROA + Covid\*Liquidity

+ Covid\*Size + Covid\*Leverage + Covid\*ROA , data = dt))

options(scipen=999)

car::vif(model2)

par(mfrow=c(2,2))

plot(model2)

bptest(model2)

shapiro.test(resid(model2))

#Predict

predictions <- predict(model1, newdata = dt)

predictions

dt$Predicted <- predictions

View(dt)

plot(as.vector(unlist(dt[3])), type = 'l', ylab = 'Value', col ="blue")

lines(as.vector(unlist(dt$Predicted)), lty = 'dotted', col='red')

install.packages("Metrics")

library(Metrics)

rmse(dt$DMS,dt$Predicted)

mean(dt$DMS)

var(dt$DMS)

#Task 6. Using Arima to predict debt maturity structure 2022

install.packages("tseries")

library("tseries")

install.packages("forecast")

library("forecast")

DMSts = ts(dt$DMS, start =c(2010, 6), end =c(2019, 12), frequency = 4)

DMSts

#Check stationary

plot(DMSts, type = "l", main = "Debt Maturity Structure",

, xlab = "Year")

adf.test(DMSts)

pp.test(DMSts)

DMSts.diff <- diff(DMSts)

plot(DMSts.diff, type = 'l')

adf.test(DMSts.diff)

pp.test(DMSts.diff)

adf.test(diff(DMSts.diff))

pp.test(diff(DMSts.diff))

acf(diff(DMSts.diff))

pacf(diff(DMSts.diff))

#Finding optimal parameter

j <- forecast:::auto.arima(diff(DMSts.diff),ic='aic',trace=TRUE)

j

accuracy(forecasted)

#Predicting

forecasted <- forecast:::forecast.Arima(j, h=4,level=c(99.5))

forecasted

X2022 <- read\_excel("2022.xlsx")

plot(as.vector(unlist(X2022[2])), type = 'l', ylab = 'Value', col ="blue")

lines(as.vector(unlist(forecasted$mean[1:4])), lty = 'dotted', col='red')

#Evaluate Matrix

X2022$forecast <- forecasted$mean[1:4]

y <- as.vector(unlist(X2022[2]))

x<- as.vector(unlist(X2022$forecast))

y

x

# Mean Absolute Error (MAE)

mae <- mean(abs(y - x))

# Mean Squared Error (MSE)

mse <- mean((y - x)^2)

# Root Mean Squared Error (RMSE)

rmse <- sqrt(mse)

# Mean Absolute Percentage Error (MAPE)

mape <- mean(abs((y - x) / y)) \* 100

mae

mse

rmse

mape

# Check white noise(p value bé hơn 0.05 thì ko có nhiễu trắng là ok)

Box.test(diff(DMSts.diff), lag = 2, type = "Ljung-Box")

shapiro.test(diff(DMSts.diff))

#Impoving arima model

t <- arima(DMSts, order =c(0,0,3))

t

k <- forecast:::forecast.Arima(t, h=4,level=c(99.5))

k

plot(as.vector(unlist(X2022[2])), type = 'l', ylab = 'Value', col ="blue")

lines(as.vector(unlist(k$mean[1:4])), lty = 'dotted', col='red')

X2022$forecast1 <- k$mean[1:4]

y1 <- as.vector(unlist(X2022[2]))

x1<- as.vector(unlist(X2022$forecast1))

y1

x1

# Mean Absolute Error (MAE)

mae1 <- mean(abs(y1 - x1))

# Mean Squared Error (MSE)

mse1 <- mean((y1 - x1)^2)

# Root Mean Squared Error (RMSE)

rmse1 <- sqrt(mse1)

# Mean Absolute Percentage Error (MAPE)

mape1 <- mean(abs((y1 - x1) / y1)) \* 100

mae1

mse1

rmse1

mape1