



Object-Oriented Programming as a Tool for Company's Overall Performance Prediction

Tufegdzc Milica*¹, Miodragovic Goran², Pravdic Predrag³

*¹Department for Information Technology, College of Applied Mechanical Engineering Trstenik, Trstenik, Serbia, E-mail: milica.tufegdzc@vtmsts.edu.rs

²Department for Information Technology, College of Applied Mechanical Engineering Trstenik, Trstenik, Serbia, E-mail: goran.miodragovic@vtmsts.edu.rs

³Department for Production Engineering, College of Applied Mechanical Engineering Trstenik, Trstenik, Serbia, E-mail: predrag.pravdic@vtmsts.edu.rs

Abstract Using and combining the capabilities of object oriented programming language Python, mathematical modelling and Balance Scorecard as a strategy performance management tool, it is possible to develop a predictive model for measuring progress towards organizational objectives. Input data, in the form of values for 4 different balanced development perspectives and overall performance from 53 organizations, are used for predictive modeling. Python programming language and its libraries such as Pandas, Matplotlib, Sklearn, Statsmodels and Thinker are used for data analysis, 2D plotting, regression, estimating statistical models and creating a Graphical User Interface. Proper scripts were written, edited and interactively tested using Spyder environment. The developed predictive model fits well into a linear model according to the value of coefficient of determination R^2 (0.887) and p values for regression coefficients in equation, which are less than 0.05. Accurate prediction of overall performance, based on the input values for learning and growth, customers and internal business processes perspectives, can be obtained by running an executable file with a simple interface, without calculating and measuring the financial perspective.

Keywords Python, regression model, overall perspective, organizational performance

Introduction

Predicting business environment and influence of certain factors is a crucial management task. New technological and science developments, rapid *increase of environmental concerns, business cycle fluctuations*, increased development of turbulent financial and business-to-business markets and globalization of business necessitate the use of several relevant techniques for accurate prediction. Analysis and research play a prominent role in understanding current and future business conditions [1]. Collection, organization, analysis, interpretation and presentation of data plays a significant role in business and helps managers make predictions and decisions about what could occur in their organization in the future. Accurate prediction model can aid managers in selecting the right combination of resources, as well as help them in making strategic decisions concerning the future, all based on the information from the past and effective reaction planning in order to achieve better organizational performance [2].

Predictive modeling is the process of developing a mathematical model that generates accurate prediction of an unknown value of a dependent variable from a set of independent variables. As a basic quantitative analytical tool, predictive modeling is used to make data-based enterprise and future predictions of organizational performances, risks and trends, by assessing variables and explaining the relationships among the variables in



the existing (recent and past) data. Quantitative data can be analyzed using specific analytical software such as R, Excel, Minitab, SAS, or SPSS [3].

Simple and multiple regression analysis are techniques for modelling and establishing a linear relationship between a dependent variable and one or more independent variables, based on the best fitted model [3, 4, 5, 6, 7]. In banking and finance literature, regression analysis is a very common method. Bank's financial performance, as a dependent variable, was measured using three indicators based on independent variables, bank's size, credit risk, asset management, operational efficiency and debt ratio. Multiple regression analysis was conducted and multiple regression model was established [7]. Different multiple regression models were obtained using SPSS computer software, as tools for forecast trading activities on the trading floor in different sectors [2]. Moderated multiple regression (MMR) was used in some management specialties such as human resources management, organizational behavior, strategy, etc [8]. Predictive modelling software such as DTREG was used in some cases, for example, to predict the bidding price using the multiple linear regression model [4].

Bearing in mind the fact that most predictive analytics software are commercial and license based, using an object-oriented programming language such as Python should be the right choice for implementing data-processing frameworks and predictive modeling of overall organizational performance, as an example. Such object-oriented approach in programming facilitates rapid development of complex applications and program reuse [9]. It provides many benefits like code readability and reusability, improved reliability and flexibility, modularity, decomposition and effective problem solving, as well as easy modifications and expansions in the future [10].

Predictive modeling of overall performance is possible using different statistical tools and techniques, among which a special attention is given to linear regression [11]. Assessing overall performance is enabled with Key Performances Indices (KPIs) arranged into 4 balanced development perspectives: finance, customers, internal business processes, learning and growth, by using Balanced Scorecard (BSC) software [12, 13].

Predictive modeling in Python was done through 3 steps: writing the script for calculating correlation coefficients and obtaining scatter plots, writing the script for best fitting regression model and its evaluation through testing statistical significance of the estimated regression parameters by t-test and F statistics, and developing a Graphical User Interface to simplify the procedure for overall performance prediction, based on the values of three perspectives, according to the results obtained for correlation coefficients after running the first script. Scripts were written, edited and interactively tested using Spyder (Scientific PYthon Development EnviRonment).

Program is suitable for fast prediction of overall performance, without calculating the financial perspective, which is sometimes a heavy task, due to collecting big amount of data. Easy running at no cost, simple interface for data entry and obtaining prediction values is possible only by running the executable file on Windows operating system.

Materials and Methods

Data were gathered from 53 organizations from Serbia in different industry branches using different methods such as interviews with responsible persons, specially prepared questionnaires and technical and economic documentation. Based on these data the values for learning and growth perspective (p_1), internal processes perspective (p_2), customer perspective (p_3) and financial perspective (p_4) were calculated in BSC Designer Light software. The values for overall perspective (p_u) were also obtained. A set of obtained values (in %), stored as Excel file, was used as input data for developing the predictive model for overall performance.

Loading and reading data from Excel file was done using `read_excel` function as 2-dimensional labeled data structure `DataFrame`, the basic data format for Pandas Python package.

Function `corr` () was used to find the pairwise correlation of all columns in the dataframe. The calculated values in IPython console are presented in Figure 1 in the form of correlation *matrix*.

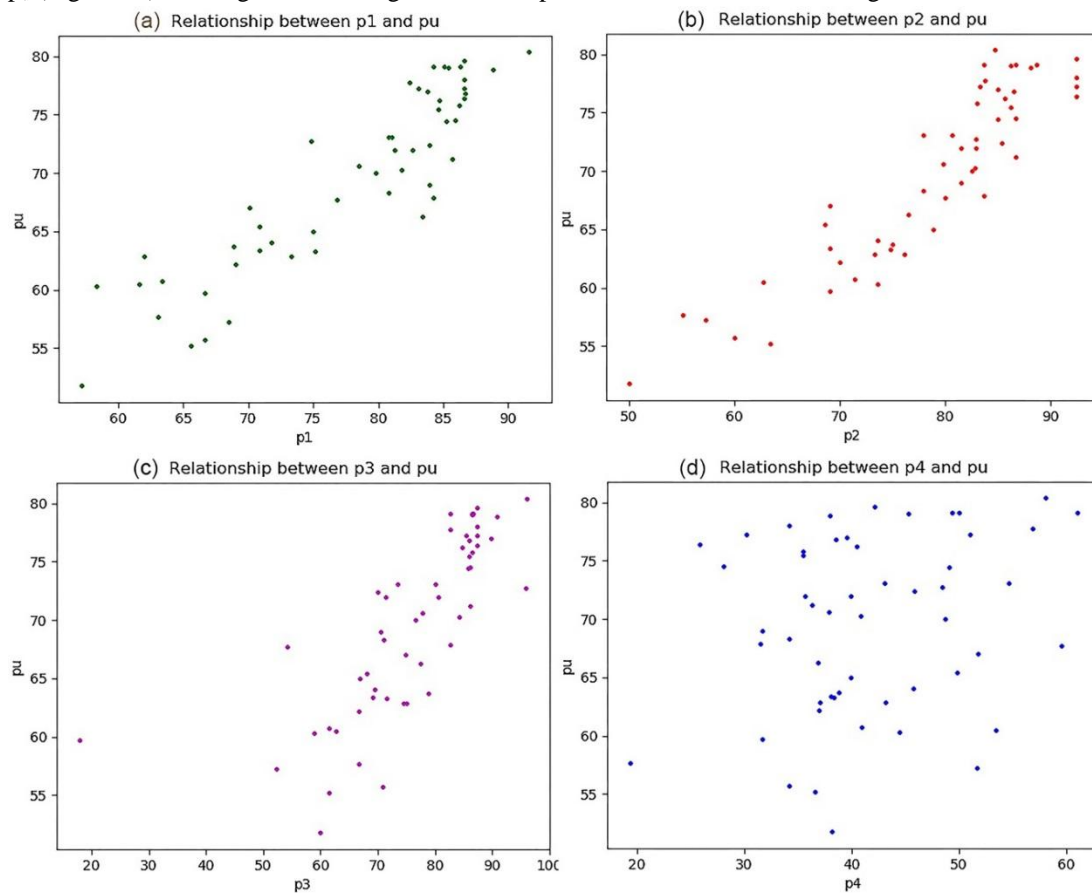


Correlation coefficients

	p1	p2	p3	p4	pu
p1	1.000000	0.873165	0.705855	0.037101	0.905050
p2	0.873165	1.000000	0.685174	0.046547	0.909121
p3	0.705855	0.685174	1.000000	0.047217	0.756790
p4	0.037101	0.046547	0.047217	1.000000	0.221554
pu	0.905050	0.909121	0.756790	0.221554	1.000000

Figure 1: Correlation matrix

Obtained values indicate that there was a strong correlation between the first 3 perspectives ($p_1 - p_3$) and p_u (values higher than 0.75). The relationship between p_4 and p_u was very low (0.221554). Object oriented plotting library Matplotlib with its submodule matplotlib.pyplot imported as plt was used for plotting scatter plots with the aim to analyze and determine the regression models (Figure 2). It is obvious that the relationship between p_4 and p_u (Figure 2d) shows great scattering and it is not possible to establish a linear regression.

Figure 2: Scatter plots for perspectives $p_1 - p_4$ and p_u in IPython console

These results, together with Correlation matrix were further used for obtaining a predictive mathematical model in the form of multiple regression equation. Overall performance (p_u) was treated as a dependent variable (predicted value), while the p_1 , p_2 and p_3 were treated as independent variables (predictors). Financial perspective was not considered as a predictor due to the fact that the relationship between p_4 and p_u was low, as it has been already stated. Mathematical expression of the proposed model is given as $p_u = \beta_0 + \beta_1 \cdot p_1 + \beta_2 \cdot p_2 + \beta_3 \cdot p_3$, where β_0 , β_1 , β_2 , β_3 represent the parameters in the model (β_0 is the intercept in regression model, $\beta_1 - \beta_3$ are the regression coefficients). To estimate the values of the parameters in the model Sklearn and Statsmodels Python packages are imported. Statsmodels provides classes and functions for the most frequently used statistical tests, such as t-test and F statistics, for example, while Sklearn with its linear_model provides the values for intercept and regression coefficients. Estimation is done using the ordinary least squares method (OLS).



Built-in graphical library called Tkinter is used to create GUI for showing regression results and providing a simple input method for p_1 , p_2 and p_3 . Predicting the value of p_u is also provided, based on obtained values for the parameters in the model. Python script was finally converted to an executable file which can be run on any other computer, requiring no source code or installed Python programming language.

Results & Discussion

The results of running the script are parameters calculation and obtaining the value for the intercept in the model. The values for the regression coefficients are shown in the form of one dimensional array (Figure 3).

```
Intercept:
9.13375599122125
Coefficients:
[0.32888168 0.34721362 0.09886882]
```

Figure 3: Values of the regression parameters in IPython console

Goodness-of-fit of the estimated regression model to the data was evaluated using a coefficient of determination (R^2) as a descriptive measure of proportion of the total variance in the dependent variable that is accounted for and explained by linear combination of the multiple independent (predictor) variables in the regression model. Its numeric value is 0.887, which indicates very good fitting to the linear model. Testing the level of statistical significance (p-value) of the estimated regression parameters was done with the aim to evaluate the goodness-of-fit. In this study, we have considered that there was a statistical significance if value of $p < 0.05$. In our case, p values were less than 0.05 for all parameters, and all of them should be considered as statistically significant. Each of the individual parameters of the regression model (the intercept and the multiple slopes) were additionally evaluated by t-test and overall test (F-statistic). The value of F statistic is large, but the significance F (Prob.) is very small, which is additional confirmation that the regression equation fits the data well. Standard errors are presented, as well as confidence intervals. The results for Omnibus and Durbin Watson test are also provided, so as the values for skew and kurtosis (Figure 4).

OLS Regression Results						
=====						
Dep. Variable:	pu	R-squared:	0.893			
Model:	OLS	Adj. R-squared:	0.887			
Method:	Least Squares	F-statistic:	136.4			
Date:	Sat, 14 Sep 2019	Prob (F-statistic):	8.79e-24			
Time:	00:58:35	Log-Likelihood:	-122.93			
No. Observations:	53	AIC:	253.9			
Df Residuals:	49	BIC:	261.7			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	9.1338	3.118	2.929	0.005	2.868	15.399
p1	0.3289	0.085	3.875	0.000	0.158	0.499
p2	0.3472	0.075	4.602	0.000	0.196	0.499
p3	0.0989	0.038	2.575	0.013	0.022	0.176
=====						
Omnibus:	0.808	Durbin-Watson:	1.591			
Prob(Omnibus):	0.668	Jarque-Bera (JB):	0.895			
Skew:	-0.255	Prob(JB):	0.639			
Kurtosis:	2.620	Cond. No.	1.20e+03			
=====						

Figure 5: OLS regression results

Simple GUI with labels and entry fields for the input values for each predictor is provided, and a simple click on the bottom gives the predicted value of overall performance as a result. The results were tested at one additional organization with the values for perspectives $p_1 = 79.51\%$, $p_2 = 95.25\%$, $p_3 = 75.36\%$, $p_u = 77.32\%$, previously calculated in BSC software (together with the value for p_4). Absolute error for predicted value, obtained in our study is 1.51%, while the value of relative error is only 0.02. Results are shown in Figure 5.



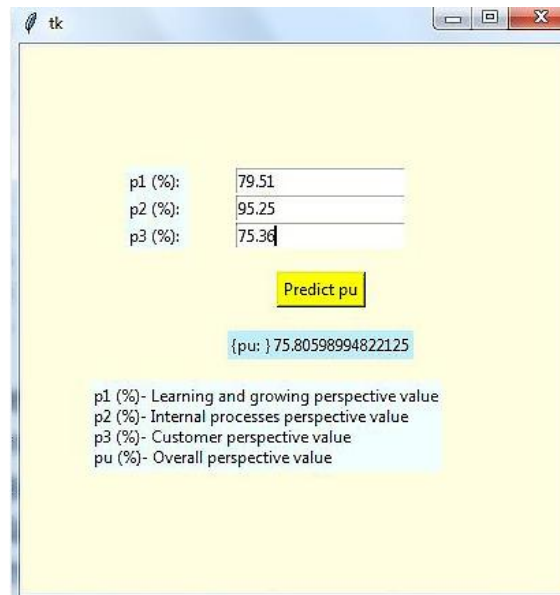


Figure 5: GUI for input data and prediction the value of p_u as a result of running the executable code

Conclusion

Presented methodology allows easy and quick calculation of predicted value with high accuracy. Special attention should be given to the fact that it is possible to make a prediction of overall performance without calculating the values for financial perspectives, which is often a heavy and time consuming task. This model should be used as a tool for specific strategic planning and decision making with the aim to track fulfillment of the organization's strategy. Implementation is easy, due to a simple GUI, which can be run as an executable file. Some improvements should be made in the future, in the sense of increasing the sample size, which will lead to higher accuracy. Obtained mathematical model will be the base for obtaining parameterized linear regression model by using neural networks.

References

- [1]. Duus, H.J. (2016). Strategic Forecasting: The Management Perspective. *Management Research Review*, 39(9):998-1015.
- [2]. Adeyemi, S.L., Aremu, M.A. (2009). The Impact of Forecasting on Strategic Planning and Decision Making: An Exploratory Study of Nigerian Stock Exchange. *African Research Review*, 3(1):179-205.
- [3]. Sema, A.K., Rafa, M.K., & Nabeel, R.K. (2016). *Enterprise Big Data Engineering, Analytics, and Management*. IGI global, Hershey, USA, 66-83
- [4]. Petrovski, A., Petruseva, S., & Zileska P.V. (2015). Multiple Linear regression model for predicting bidding price. *Technics technologies education management*, 10(3):386-393.
- [5]. Rawlings, J.O., Pantula, S.G., & Dickey, D.A. (1998). *Applied Regression Analysis, A Research Tool*. 2nd Ed., Springer-Verlag, New York, Inc., 9-10
- [6]. Anghelache, C., Anghel, M.G., Prodan, L., Sacală, C., & Popovici, M. (2014). Multiple Linear Regression Model Used in Economic Analyses. *Revista Română de Statistică - Supliment*, 10:120 - 127.
- [7]. Nataraja, N.S., Nagaraja, R.C., & Ganesh, L. (2018). Financial performance of private commercial banks in India: multiple regression analysis. *Academy of Accounting and Financial Studies Journal*, 22(2): 1-12.
- [8]. Herman, A. (1995). Statistical Power with Moderated Multiple Regression in Management Research. *Journal of Management*, 21(6):1141-1158.
- [9]. Asagba, P.O., & Ogheneovo, E.E. (2008). A Comparative Analysis of Structured and Object-Oriented Programming Methods. *Journal of Applied Sciences and Environmental Management*, 12(4):41 - 46.



- [10]. Jassim, F., & Altaani, F. (2013). Statistical Approach for Predicting Factors of Mood Method for Object Oriented. *IJCSI International Journal of Computer Science Issues*, 10(1): 589-593.
- [11]. Rodríguez, R.R, Saiz, A.J.J., Bas, O.A., & Sáez, V.M.J. (2010). Identifying relationships between key performance indicators. In: *4th International Conference on Industrial Engineering and Industrial Management XIV Congreso de Ingeniería de Organización Donostia - San Sebastián*. September 8th - 10th, 151-159.
- [12]. Haddadi, F., & Yaghoobi, T. (2014). Key indicators for organizational performance measurement. *Management Science Letters* 4, 2021–2030.
- [13]. Kádárová, J., Durkáčová, M., Kalafusová, L. (2014). Balanced Scorecard as an issue taught in the field of industrial engineering. *Procedia - Social and Behavioral Sciences*, 143:174 – 179.

