**PROFIT PREDICTION FOR**

**50 COMPANIES**

**Organization:** Exposys data labs

**Naresh S**

*3rd year student of BTech Artificial Intelligence and Data Science*

*Easwari engineering college , Chennai.*

*ainaresh1511@gmail.com*

Keywords: Linear Regression, Lasso Regression , Elastic Net Regression, Ridge Regression

**1.Abstract**

The ability to predict profit is impossible without a computerized system as many factors must be taken into consideration. In this program, machine learning algorithms are used to predict profit from R&D costs, administration costs, and marketing expenses. Four machine learning algorithms are used in this analysis, such as linear regression, ridge regression, lasso regression, and elastic net regression, to derive a new prediction that is more reliable than a single algorithm.

**2.Introduction**

Data is produced everywhere in today's world, like when traveling to different locations (GPS data), browsing the internet (internet history), storing pictures, etc. In order to provide a personalized environment to the user, these information's are being used. The challenge is that these data are quite large, and they cannot be processed by a single person or even a team because their sources of production (if a mobile device is turned on then data is generated from that device) make them quite challenging to process. In order to provide users with what they want, Machine Learning makes use of all these data. A core concept of Machine Learning is used to predict the profit of a company, since determining or predicting the profit of any company has become quite challenging in recent years. As many factors affect a company's profit, including R&D costs, administration, marketing, and company standards, the profit of a company is affected by a number Increasing factors affect a company's profit, making things unpredictable for an average person. individual. By analyzing the history of the companies, such as their previous profit record, administration costs, a model is developed that recognizes patterns based on the factors that affect profit, so that profit can be predicted more accurately.

**3.Methodology**

**3.1 Machine Learning**

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

**3.2 Linear Regression**

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

**3.3 Ridge Regression**

Ridge regression is a method of estimating the coefficients of multiple-regression models in scenarios where the independent variables are highly correlated. It has been used in many fields including econometrics, chemistry, and engineering.

**3.4 Elastic Net Regression**

Elastic net is a penalized linear regression model that includes both the L1 and L2 penalties during training. Using the terminology from “The Elements of Statistical Learning,” a hyperparameter “alpha” is provided to assign how much weight is given to each of the L1 and L2 penalties.

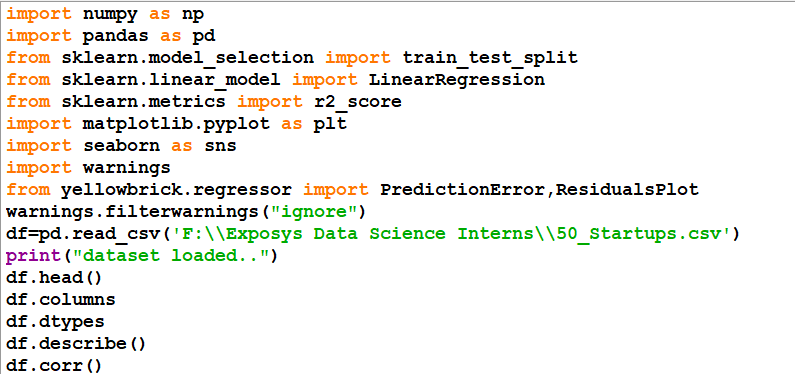
**3.5 Lasso Regression**

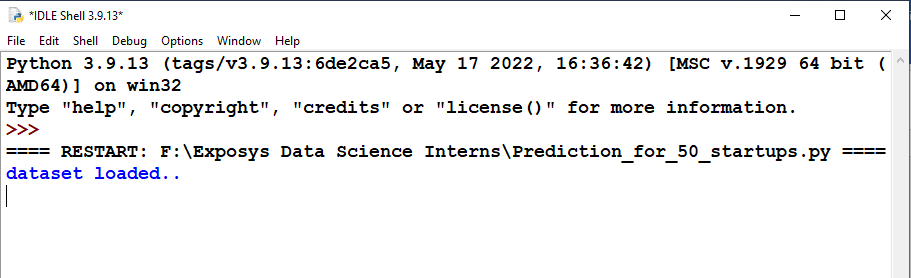
In statistics and machine learning, lasso (least absolute shrinkage and selection operator; also Lasso or LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model.

**4. Proposed System and Implementation**

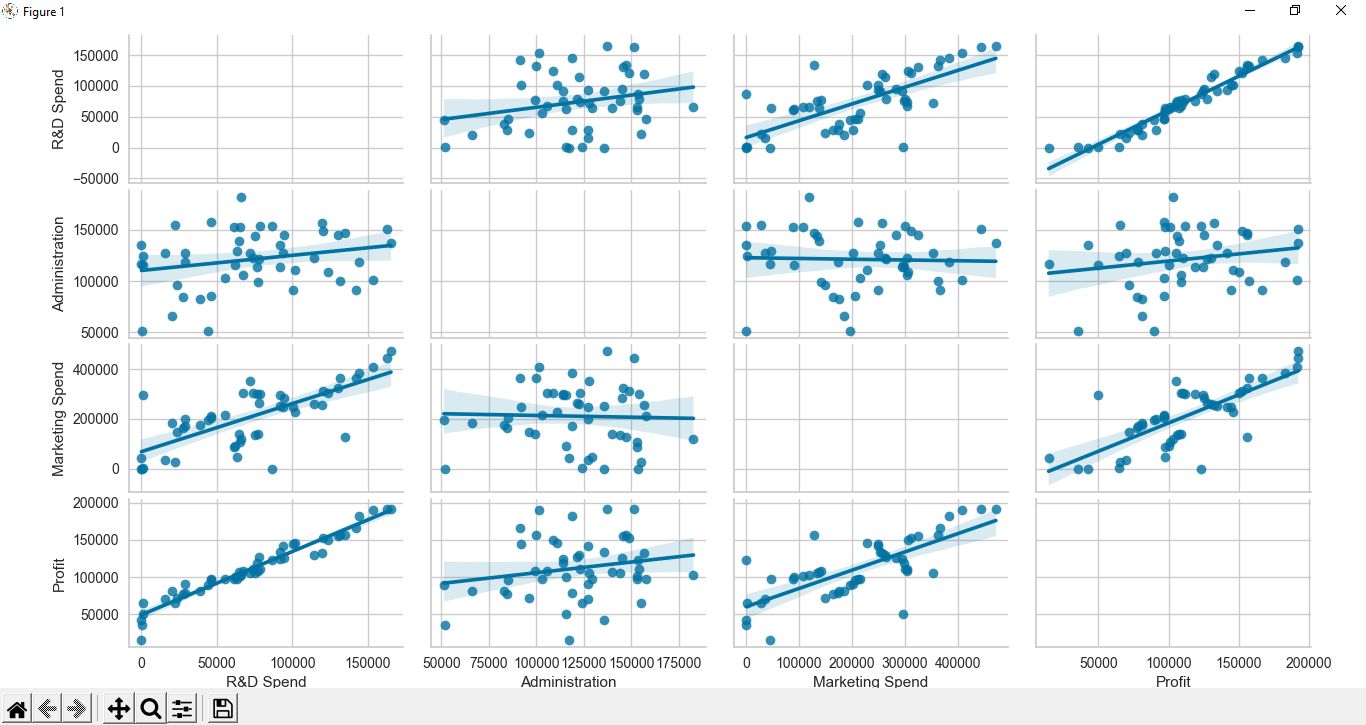
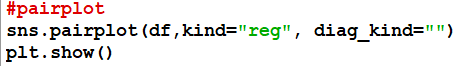
The main intention is to predict the value of the dependent variable i.e., the value of the profit of the company based on the data of the company over the previous years. So, from all the techniques used before for the prediction of profit an average from all those predicted values of the dependent variable is computed and made as the predicted dependent variable.

**4.1 Dataset implementation**

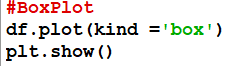


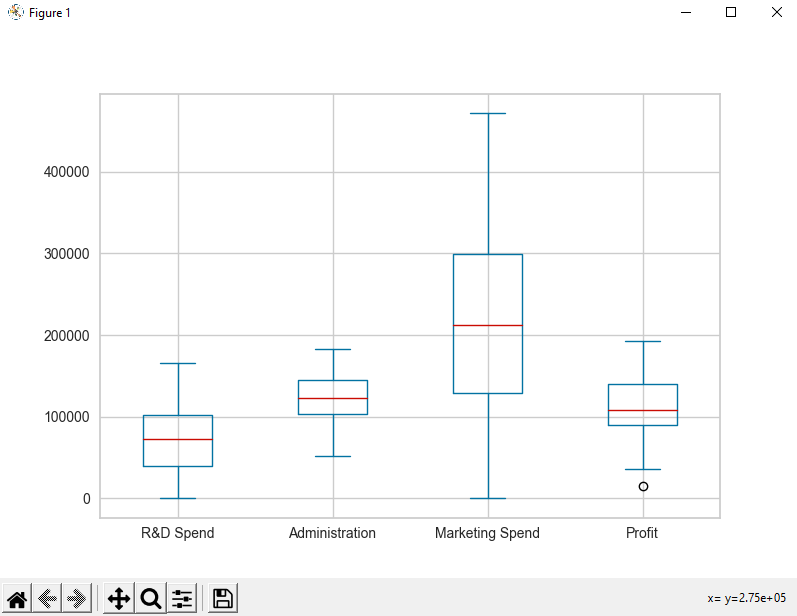


**4.2 Pair plot visualization**

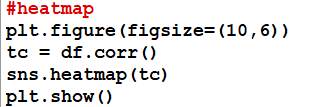


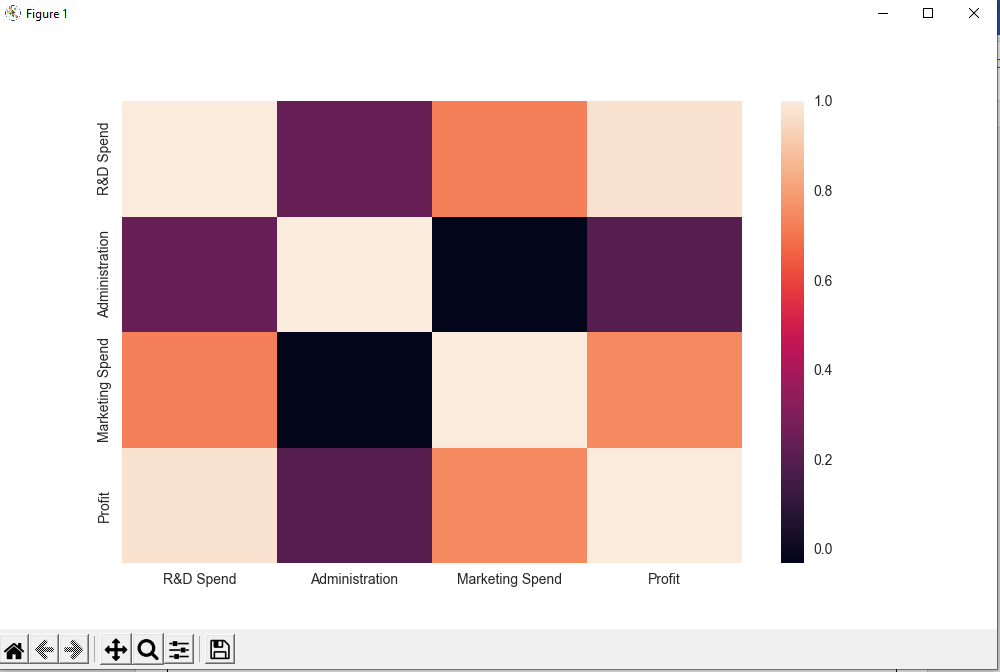
**4.3 Box Plot Visualization**



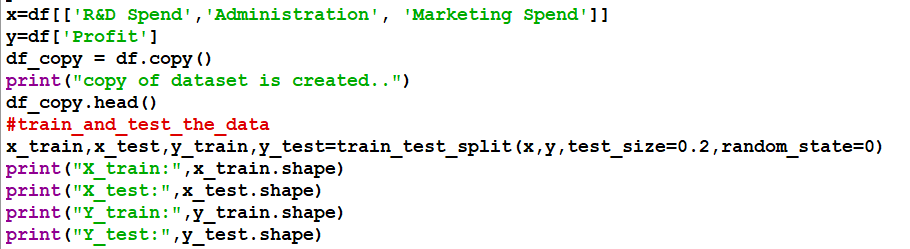
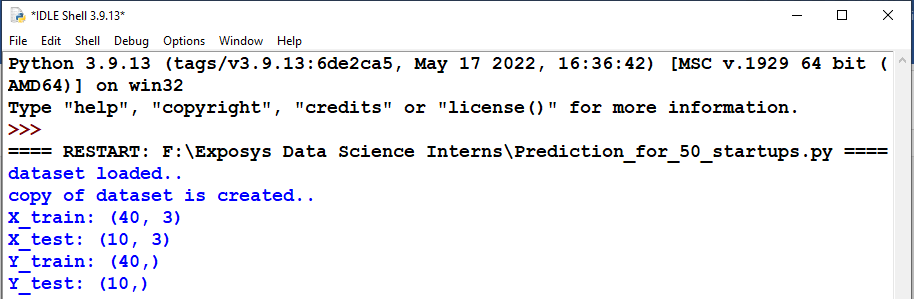


**4.4 Correlation Heatmap**



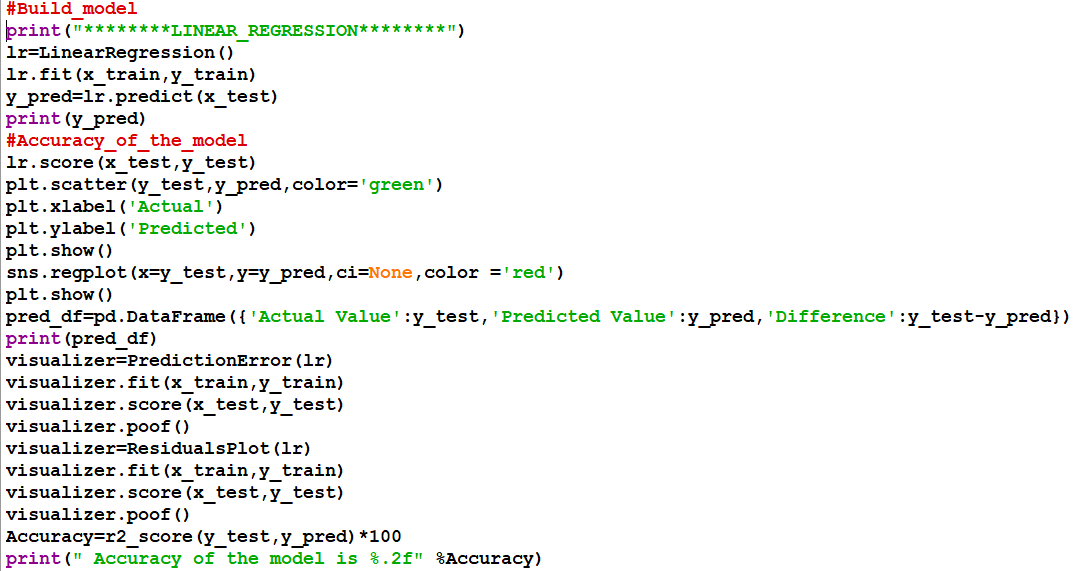


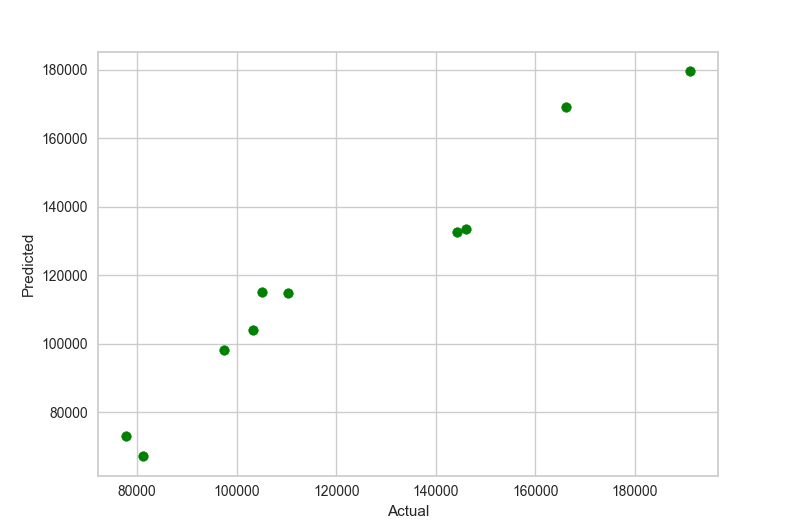
**4.5 Train and test data**

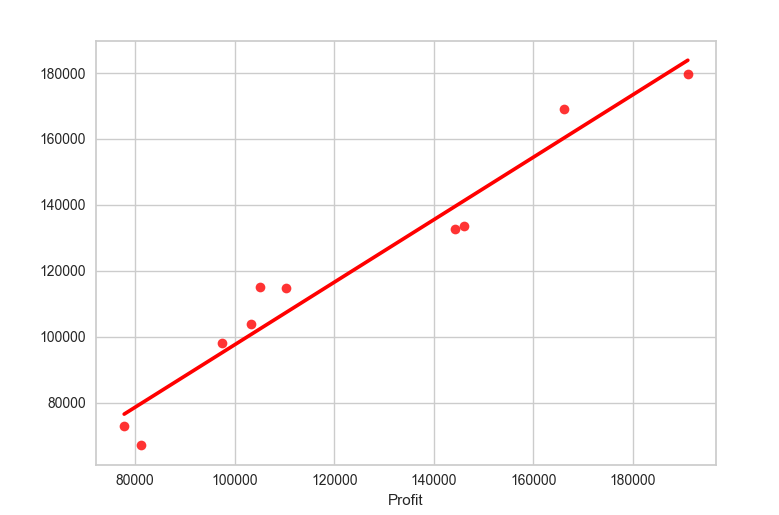


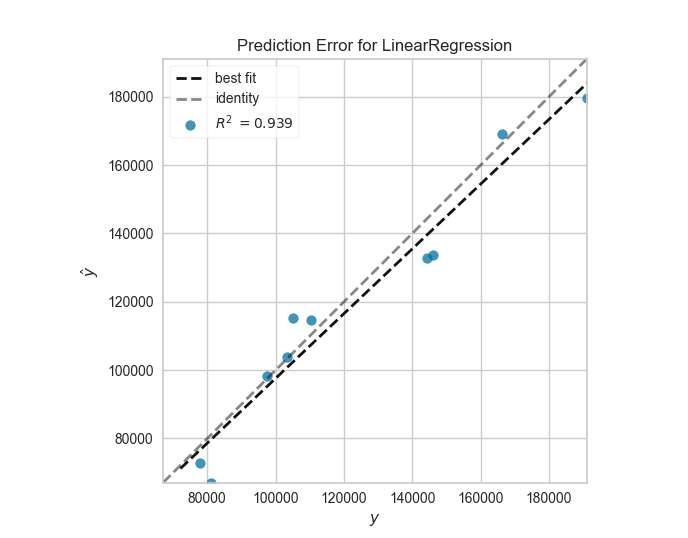
**4.6 Building Model**

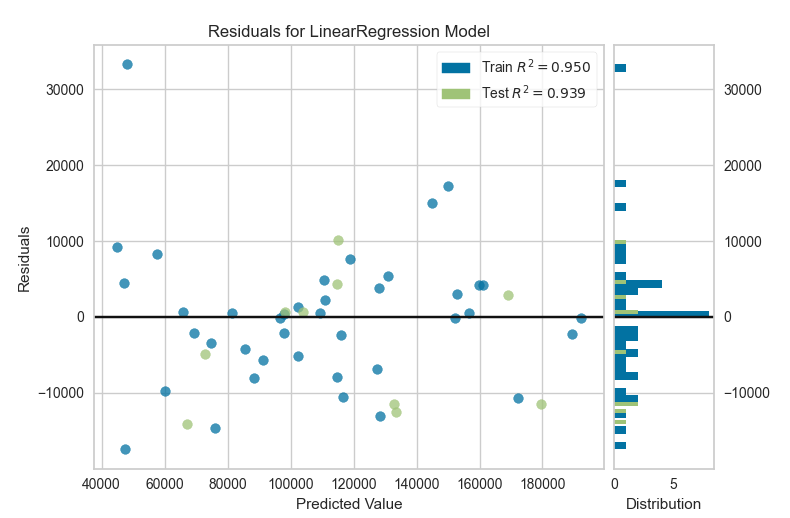
**4.6.1 Linear Regression**

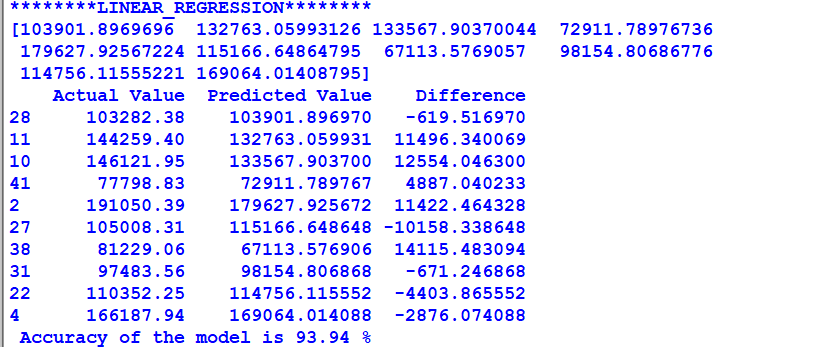




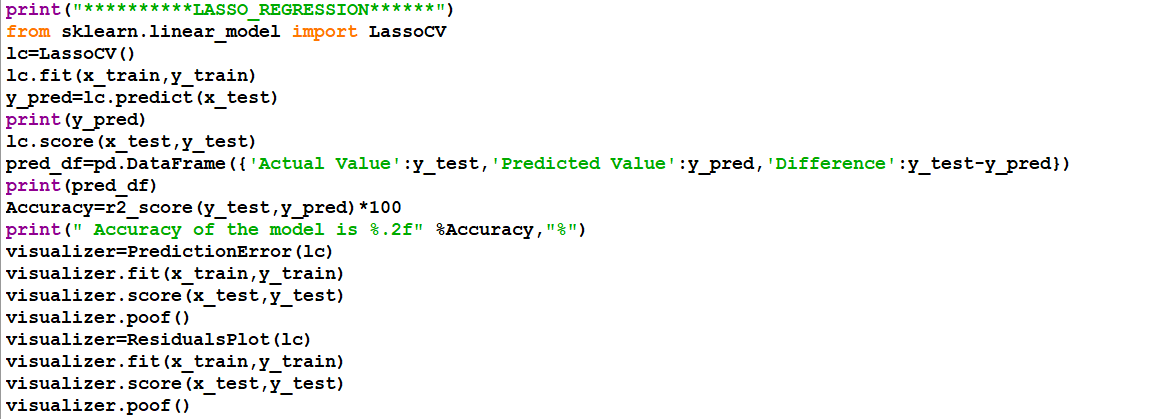


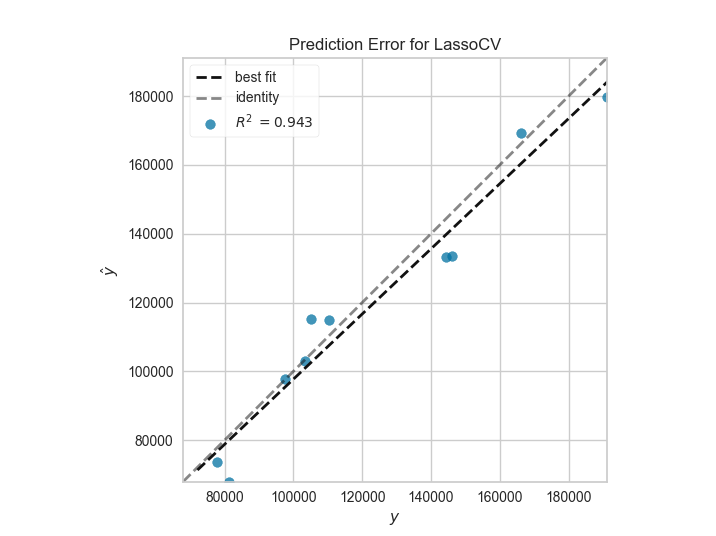


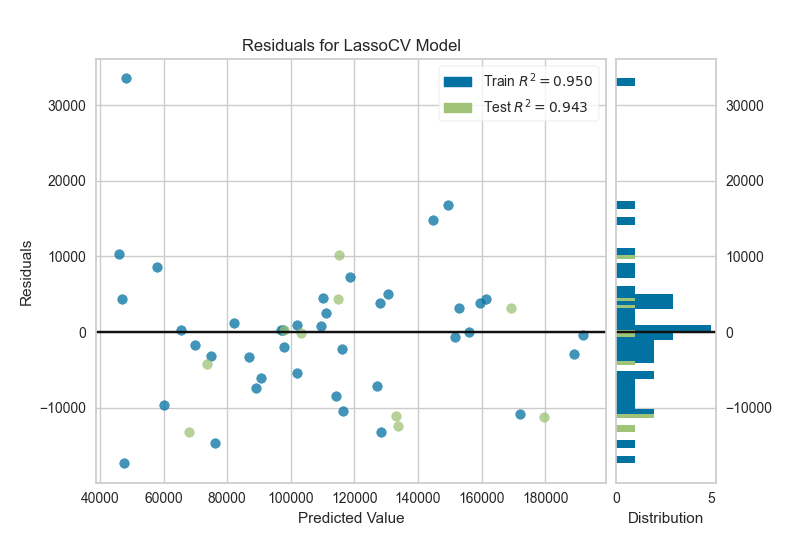


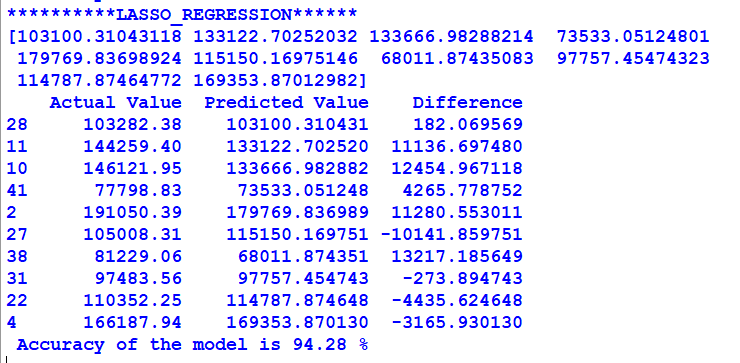


**4.6.2 Lasso Regression**

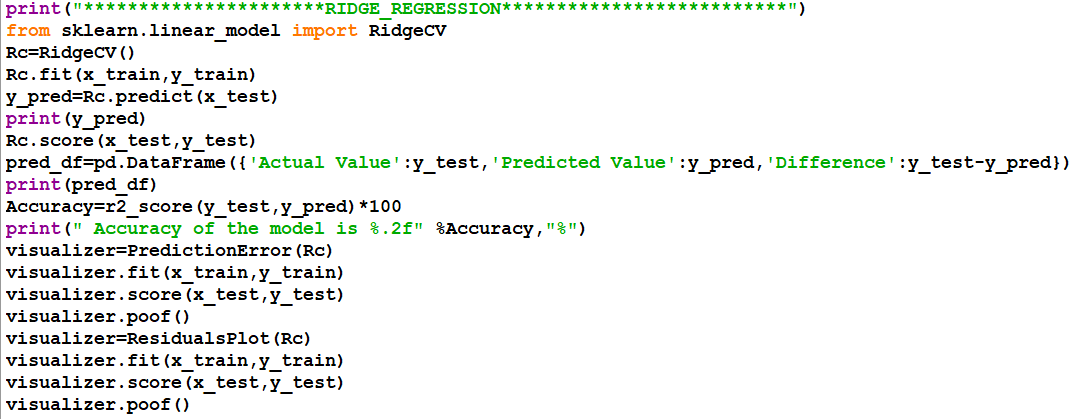


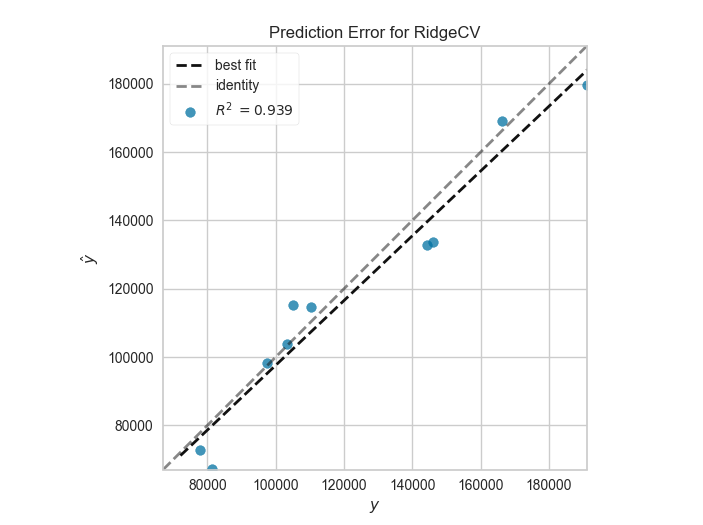


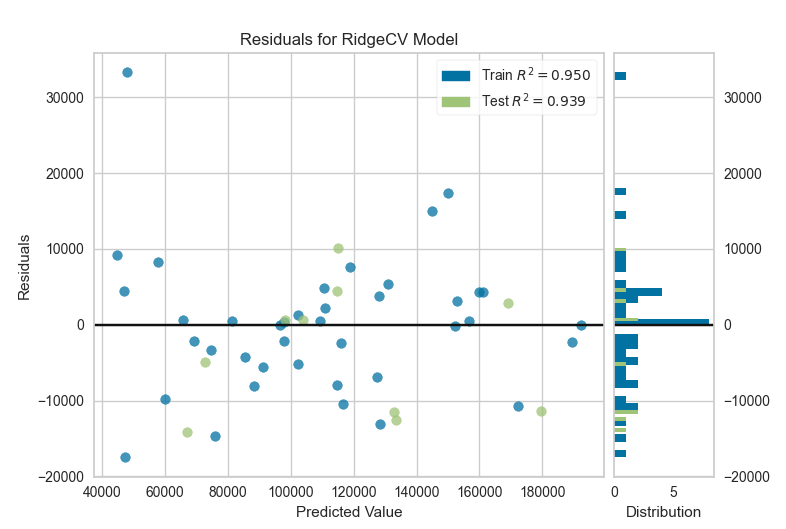


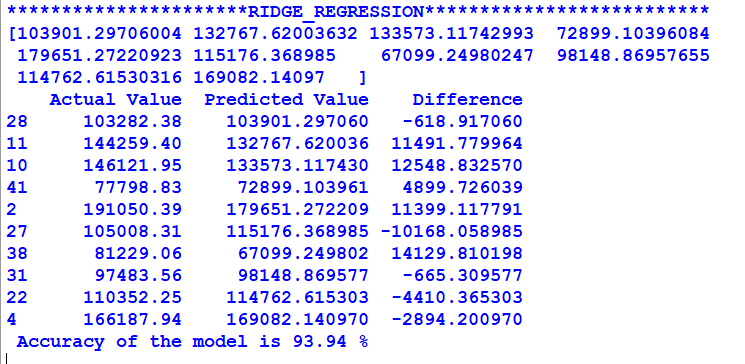


**4.6.3 Ridge Regression**

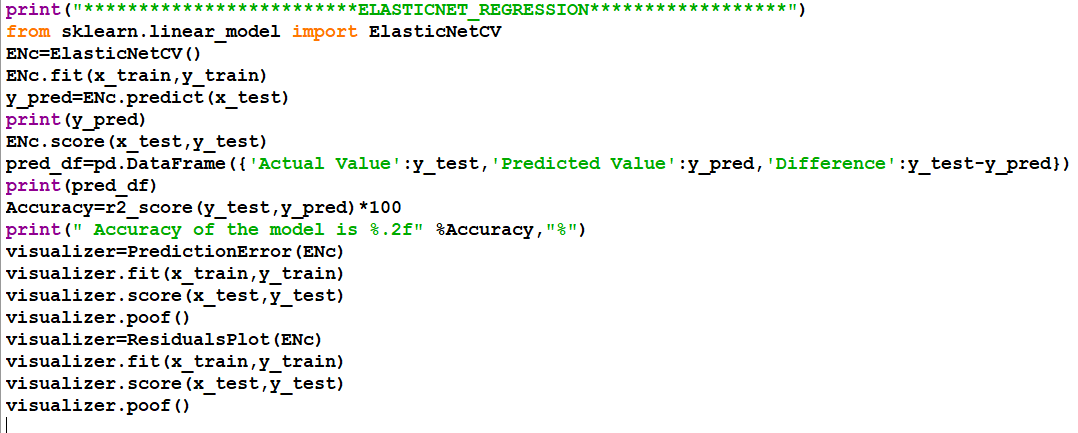


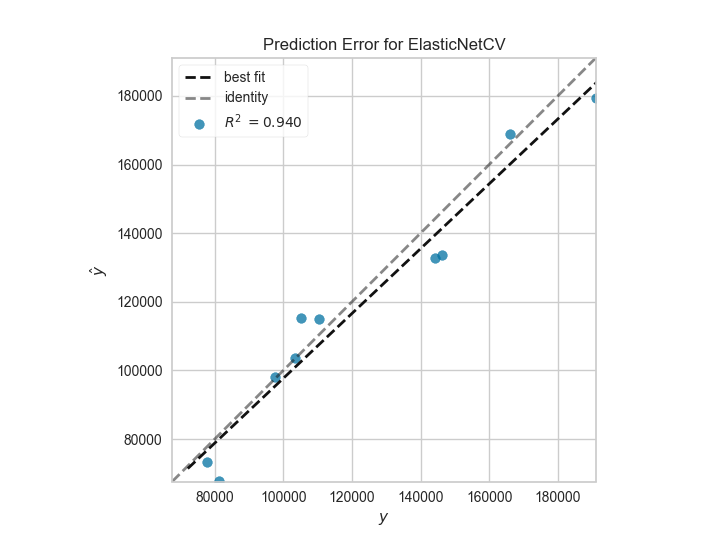


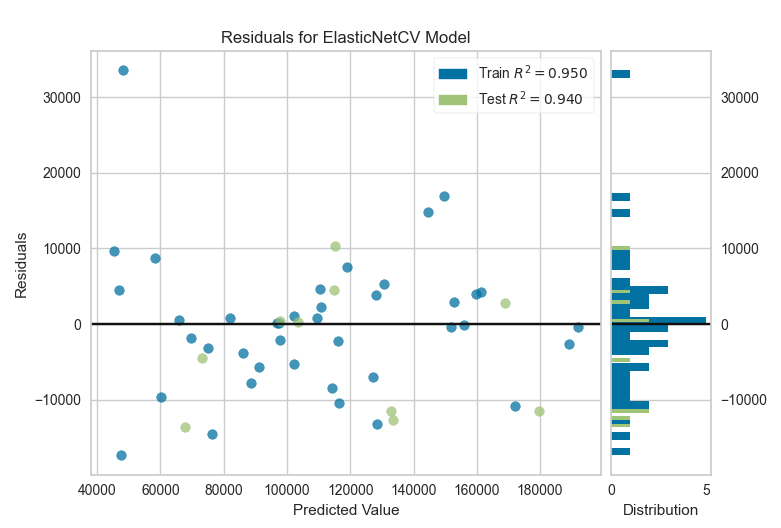


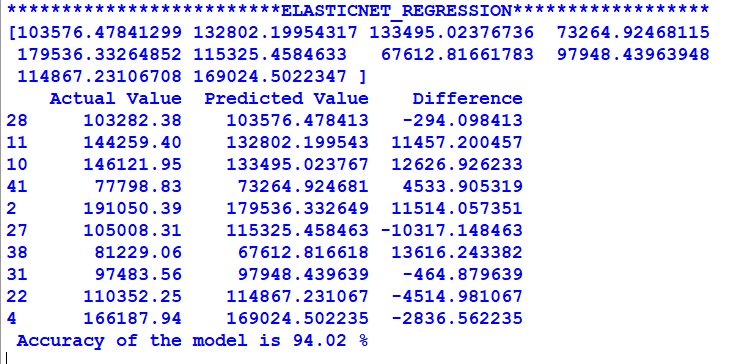


**4.6.4 Elastic Net Regression**









**5.Conclusion**

In the above model trained dataset, machine learning algorithms such as linear regression, ridge regression, lasso regression, and elastic net regression were applied. For this ML model that can predict a company's profit value based on its R&D Spend, Administration Cost, and Marketing Spend, Lasso regression gave the highest accuracy among these four algorithms.

**6.References**

[1] Understanding Multiple Regression (The fundamental basis)

[2] https://towardsdatascience.com/understanding-multiple-regression-249b16bde83e.

[3] Support Vector Regression Tutorial for Machine Learning,

[4] https://www.analyticsvidhya.com/blog/2020/03/support-vector-regression-tutorial-for-machinelearning/ .

[5] An introduction to support vector regression, https://towardsdatascience.com/an-introduction-tosupport-vector-regression-svr-a3ebc1672c2

[6] How Random Forest works, https://builtin.com/data-science/random-forest-algorithm .

[7] Understanding Random Forest (How the Algorithm Works and Why it Is So Effective) and Decision

Tree, https://towardsdatascience.com/understanding-random-forest-58381e0602d2 .

[8] “The Politics”, by Aristotle in 1877.

[9] Decision Making, the right way to use the wisdom of crowds, https://hbr.org/2018/12/the-right-wayto-use-the-wisdom-of-crowds .

[10] What makes Python the best programming language for machine learning and the best programming language for AI? https://steelkiwi.com/blog/python-for-ai-and-machine-learning/ .

[11] Fabian Pedregosa; Gaël Varoquaux; Alexandre Gramfort; Vincent Michel; Bertrand Thirion; Olivier Grisel; Mathieu Blondel; Peter Prettenhofer; Ron Weiss; Vincent Dubourg; Jake Vanderplas; Alexandre Passos; David Cournapeau; Matthieu Perrot; Édouard Duchesnay (2011).

[12] "Scikit-learn: Machine Learning in Python" http://jmlr.org/papers/v12/pedregosa11a.html . Journal of

Machine Learning Research. 12: 2825–2830.