Prediction Using Regression Analysis

Shantanu Sarkar[1], Anuj Vaijapurkar[2], VimalKumar Bhardwaj[3],Swarnalatha P[4]

[1,2,3]School of Computer Science, VIT University, Vellore
[4] Assistant Professor, School of Computer Science, VIT University, Vellore

---***---

Abstract – *Regression analysis is very important tool in statistics and it is widely used for prediction. Regression analysis can be done on any dataset and correctness of the model can be checked using regression coefficient. This paper is based on the tool created for regression analysis which is based on simple linear regression model. This tool provides output which are different results a person needs for regression analysis for example t test value, F-test value etc.*

This project is made for the general public who would like to utilize the functionalities of economic features. This is a Simplification on existing Tools. This tool doesn't require any prior knowledge of statistics and output is given such that anyone can easily understand it.

Keywords: *Regression, Linear Regression, Analysis, Python, Tools*

1. INTRODUCTION

Regression analysis is a statistical procedure for evaluating the relationship among variables. It consists of techniques to model and analyse variables. Regression analysis is used to understand the variation in the dependent variables when there is a variation in any of the independent variables, keeping the other independent variables constant. Regression analysis is broadly utilized for prediction and estimating, where its utilization has considerable cover with the field of machine learning. Regression analysis is likewise used to comprehend which among the autonomous factors are identified with the needy variable, and to investigate the types of these connections. In confined conditions, regression analysis can be utilized to derive causal connections between the independent and dependent variables. However this can prompt to deceptions or false connections, so staying alert is advisable. Numerous systems for performing regression analysis have been produced. The execution of regression analysis techniques relies on upon the type of the information producing procedure, and how it identifies with the relapse approach being utilized. Since the genuine type of the information creating procedure is for the most part not known, regression analysis regularly depends to some degree on making assumptions about this procedure. These assumptions are testable sometimes if an adequate amount of information is accessible. Regression models are frequently valuable notwithstanding when the suppositions are violated, in spite of the fact that they may not perform ideally. In numerous applications, particularly with little impacts or inquiries of causality in view of observational information, regression analysis can give deceiving outcomes.

In our system we have used the Simple Linear Regression Model (SLRM), Simple linear regression is similar to a linear regression model with an additional single explanatory variable. That is, the simple linear regression model consists of a two-dimensional system with a dependent and an independent variable, by finding linear function that, as precisely as probable predicts values of variables which are dependent as a function of values which are independent. Linear regression actualizes a statistical model that, when connections between the dependent and the independent variables are linear, shows ideal results.

2. LITERATURE REVIEW

There are many tools available for the regression analysis for example excel has some extension tools which allows us to do regression analyisis in the excel itself. Then there are statistical software like SPSS or NCSS which gives this option.

The facility provided in the excel requires latest excel package and we need to download analysis toolkit to run regression analysis in the excel. Excel provides an option to select dependant and independent variable after providing the data and it gives the result in form of statistical reports which includes anova test report and general terms related with the regression analysis like R square value, Multiple F value etc. It also plots the given point on the 2D surface and gives the approximate regression analysis line.

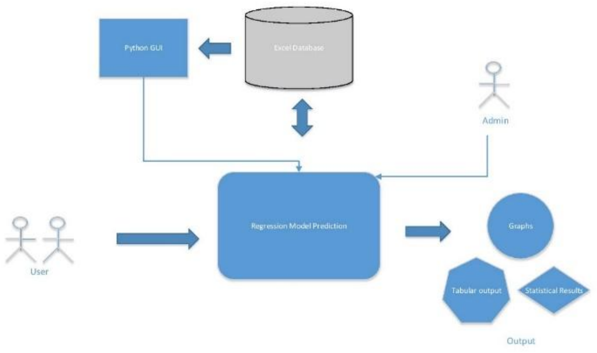
Regression Analysis alludes to a gathering of systems for considering the connections among at least two variables in view of an example. NCSS makes it simple to run either a basic direct regression analysis or a complex regression analysis, and for an assortment of responses. NCSS has present day graphical and statistical devices for concentrating on residuals, multicollinearity, decency of-fit, model estimation, relapse diagnostics, subset determination, investigation of variance, and numerous different perspectives that are particular to kind of regression being performed. Manufactured by NCSS, LLC, it is a statistical package which was built in 1981 by Jerry Hintze. NCSS, LLC has some expertise in giving measurable software for statistical analysis to businesses, academics institutions, and researchers. It likewise delivers PASS Sample Size Software which is utilized as a part of logical study arranging and assessment.

Statistical Package for Social Sciences (SPSS) is a software package utilized for statistical investigation. SPSS is a generally utilized program for factual investigation as a part of social science. It is likewise utilized by health researchers, marketing organizations, government, market researchers and others. The first SPSS manual (1970) has been portrayed as one of "humanism's most compelling books" for permitting common analysts to do their own measurable analysis. Adding statistical analysis, information administration and information documentation are components of the base programming software. The base software consists of the following statistics: Bivariate Statistics, Prediction for acknowledging groups, prediction for statistical outcomes, Descriptive Statistics. SPSS has a lot of features which can be accessed by pull down menus or it can be customized with a restrictive 4GL command syntax.

3. METHODOLOGY

Our software uses web based GUI which support latest browsers. User can use excel file to store the data on which he wish to apply regression analysis. Since data handling is easy in excel for both user and at the server's end we have included this facility. To process the data at the server's end we use python compiler since it allows us to manipulate the data easily and efficiently also python programs takes less memory and high order matrix

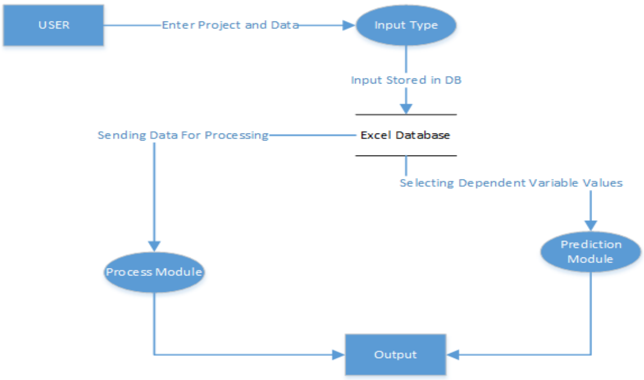
calculation can be easily done in python which is base for the results our software produces. Overall diagram of how our system works is given below:-



As shown above user gives his dataset into our system through web GUI the file format should be .xlxs. This dataset is stored into server's database and for further calculation this data is used. Notification for successful file upload or error message if any is shown on the GUI by which user can understand if the correct data is stored in the database or not. Up till this user can modify his dataset by reuploading the dataset file into server.

For manipulation of above data we use python program which calculates the output values for user. Python compiler reads the data from the database in terms of matrices since, mathematical operations on matrices is easy. To store these matrices in python we use list datatype which is built in datatype provided by python compiler. Once compiler get these values server program run's the python code which return the output values. These values are stored in the database for future calculations along with the intermediate results.

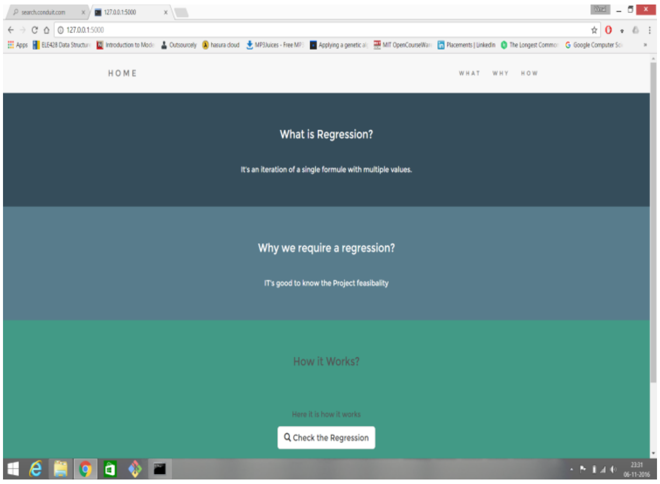
Later on these results are shown on the web GUI in terms of graph and statistical report for user this is the final output by which he can make predictions.



Above diagram shows data flow in the system. So first data is transferred into database by the user further these data is given as input to the regression predictor. This module manipulates the data and computes the output which is shown to the user on the screen.

4. IMPLEMENTATION

To implement this project we require following softwares
At the server end
Server:- Python Flask
Manupulation code:- Python language
Database:- Excel
GUI:-HTML,CSS,JavaScripts
Client end:- Since the GUI is base on HTML,CSS and JavaScripts user need lastest browser to access our tool e. g Firefox, Google Crome.
This section shows the screen shots of the above system implemented in windows environment and the description about the screen



What is Regression?

It's an iteration of a single formule with multiple values.

Why we require a regression?

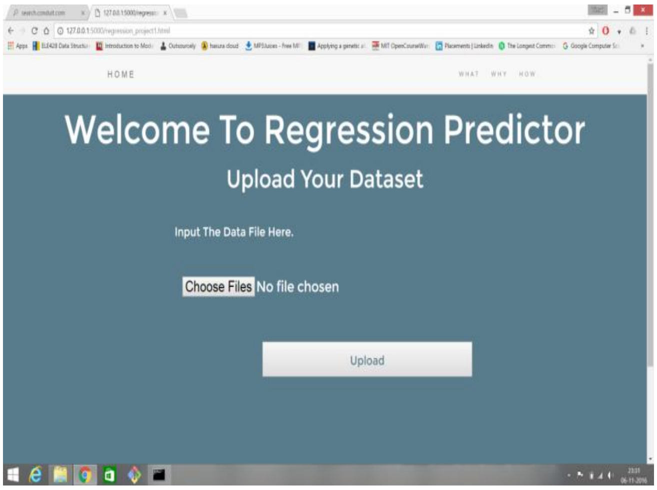
IT's good to know the Project feasibality

How it Works?

Here it is how it works

Check the Regression

This is introductory screen which tells user about regression analysis and it's uses also how it works we have included this information so that any person who doesn't know about statistics could also use our tool for his benefits. This screen provides the button which will navigate the user to the next page where user can give input into the system and get the result.



Welcome To Regression Predictor

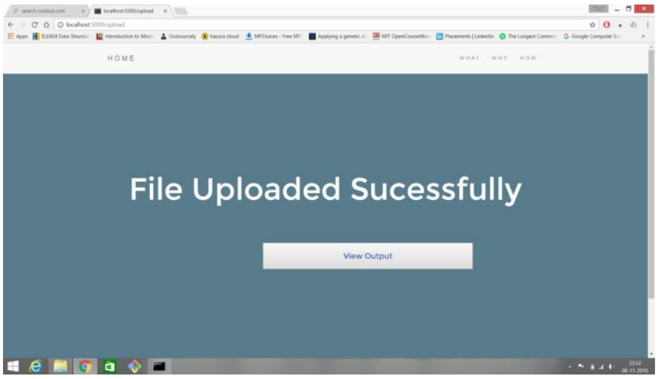
Upload Your Dataset

Input The Data File Here.

Choose Files No file chosen

Upload

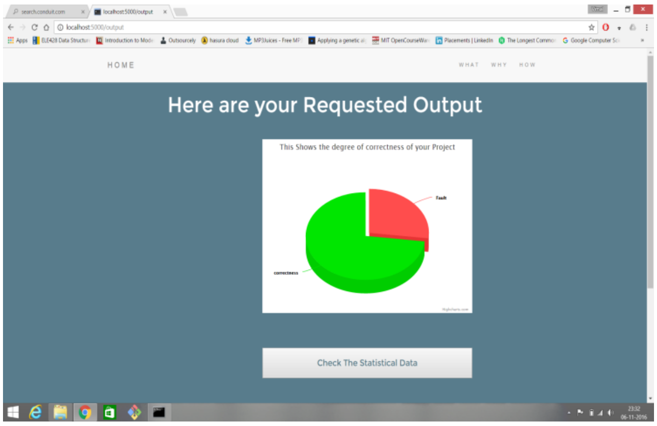
This window allows user to upload his dataset into system. This dataset will be considered for regression analysis.



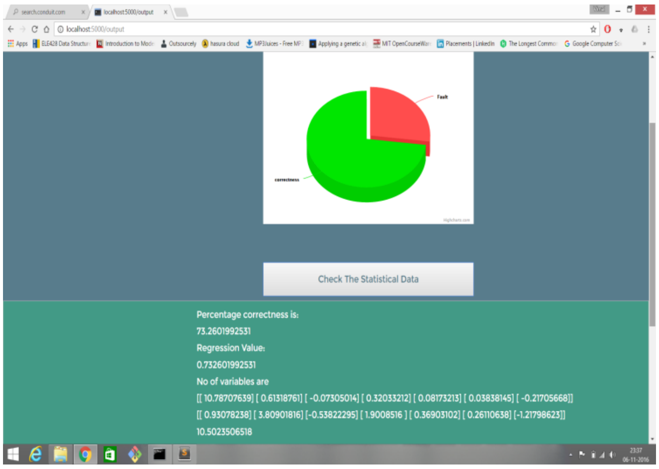
File Uploaded Sucessfully

View Output

Once the file is successfully uploaded user will see above window. If some error occurs while upload dataset user will get an error message here. Also gives option to user to upload the file again. This window will allow user to go on output window.


HOME
WHAT
WHY
HOW
Here are your Requested Output
This Shows the degree of correctness of your Project
Fault
correctness
Check The Statistical Data

This window shows the correctness of the model for the given data. Our system provides the function where user can see statistical output also if he wishes to by clicking on the button provided.



```
Percentage correctness is:
73.2601992531
Regression Value:
0.732601992531
No of variables are
[[ 10.78707639] [ 0.61318761] [ -0.07305014] [ 0.32033212] [ 0.08173213] [ 0.03838145] [ -0.21705668]]
[[ 0.93078238] [ 3.80901816] [-0.53822295] [ 1.9008516 ] [ 0.36903102] [ 0.26110638] [-1.21798623]]
10.5023506518
```

This window shows the statistical values for the tests.

5. FUTURE WORK

Currently this system works on simple linear regression model (SLRM) so in future this concept could be applied on multiple linear regression model. Also regression analysis is dependent upon six assumptions which includes tests for multi co linearity etc those tests are not included in the current version of the system. Also prediction model could be added in the same system.

6. CONCLUSION

This paper gives the new method which is user friendly, free of costs and easy to operate. We finally conclude since this system is based on web based GUI it is easy to use than other available software packages. Also with the different dataset the system works correctly.

7. REFERENCES

[1]Chatterjee, Samprit, and Bertram Price. *Regression Analysis by Example*. New York: Wiley, 1977. Print.

[2]Docs. Python. org,. "13. 1. Csv — CSV File Reading And Writing — Python 2. 7. 11 Documentation". N. p. , 2016. Web. 26 Jan. 2016.

[3]Site Point,. "Using Python To Parse Spreadsheet Data". N. p. , 2015. Web. 26 Jan. 2016.

[4]Wikipedia,. "Regression Analysis". N. p. , 2016. Web. 26 Jan. 2016.

[5]Cameron, Adrian Colin, and P. K Trivedi. *Regression Analysis Of Count Data*. Print.

[6]Gunst, Richard F, and Robert L Mason. *Regression Analysis And Its Application*. New York: M. Dekker, 1980. Print.

[7]Reliawiki. org,. "Multiple Linear Regression Analysis - Reliawiki". N. p. , 2016. Web. 26 Jan. 2016.

[8]Wikipedia,. "Linear Regression". N. p. , 2016. Web. 26 Jan. 2016.

[9]"Regression Analysis Software | Regression Tools | NCSS Software". *Ncss. com*. N. p. , 2016. Web. 6 Nov. 2016.

[10]"SPSS". *En. wikipedia. org*. N. p. , 2016. Web. 6 Nov. 2016.

[11]"Sample Regression Analysis - Wikihow". *Wikihow. com*. N. p. , 2016. Web. 6 Nov. 2016.