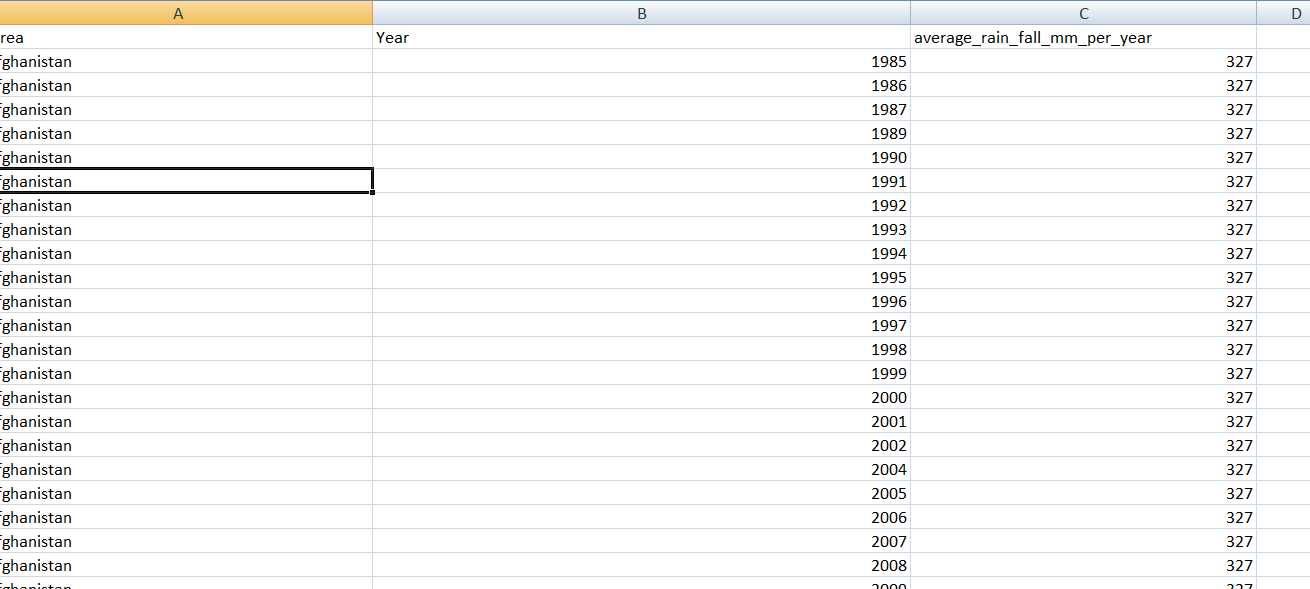
1. Consider the following rain fall prediction dataset and apply the following

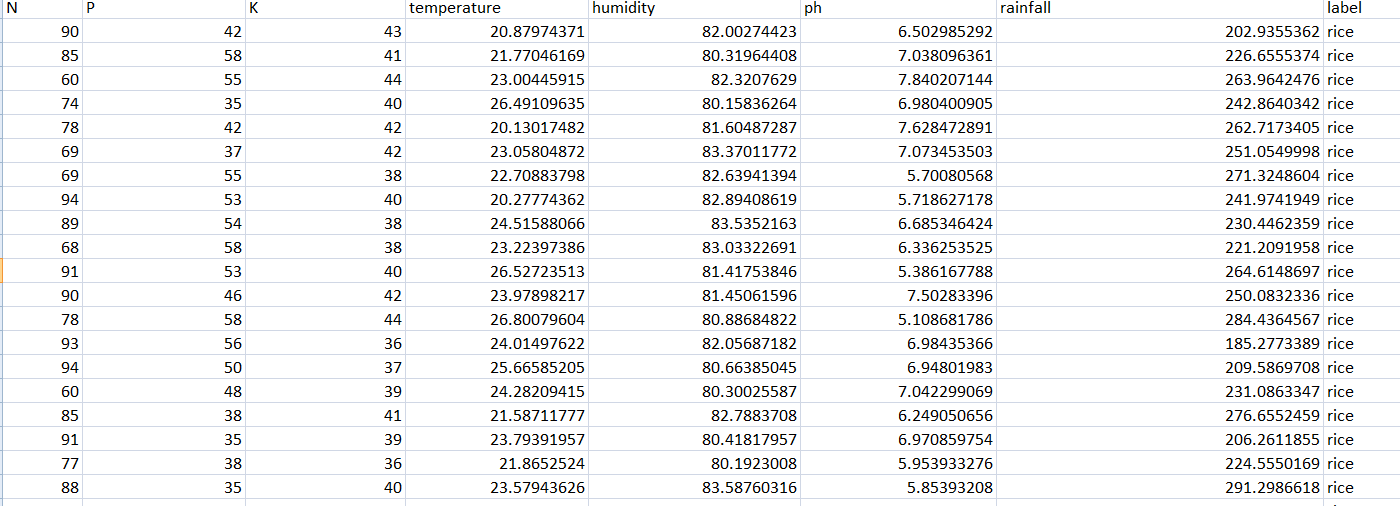
|  |  |  |
| --- | --- | --- |
| Area | Year | average\_rain\_fall\_mm\_per\_year |
| Afghanistan | 1985 | 327 |
| Afghanistan | 1985 |  |



* Find the statistical values
* Plot the data
* Apply label encoding on area
* Build the Lasso Regularized regression(Write the model formulae) Evaluate the R2 and Errors

1. Consider the Crop Recommendation Dataset.csv

This dataset was build by augmenting datasets of rainfall, climate and fertilizer data available for India.



Data fields

* N - ratio of Nitrogen content in soil
* P - ratio of Phosphorous content in soil
* K - ratio of Potassium content in soil
* temperature - temperature in degree Celsius
* humidity - relative humidity in %
* ph - ph value of the soil
* rainfall - rainfall in mm
* label(rice,potato…)

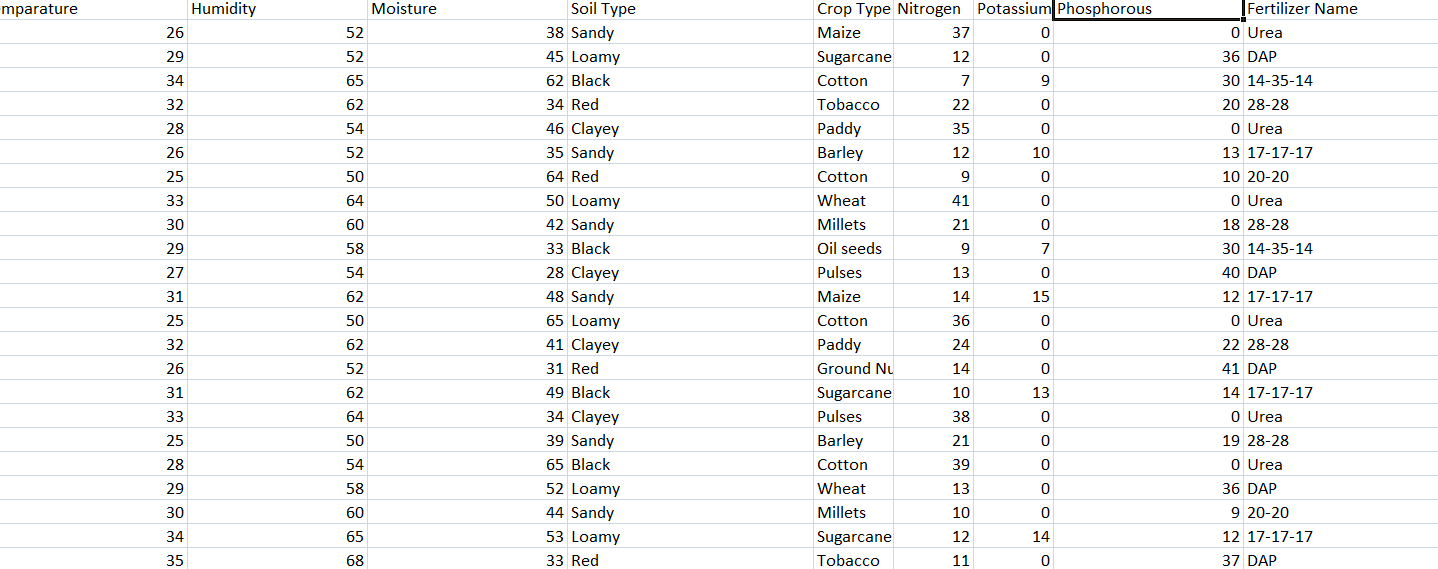
1. Find the statistical values
2. Visualize relationship using scatter plot
3. Build Logistic Regression Model and Evaluate metrics
4. Consider the Crop Recommendation Dataset.csv

This dataset was build by augmenting datasets of rainfall, climate and fertilizer data available for India.

Data fields

* N - ratio of Nitrogen content in soil
* P - ratio of Phosphorous content in soil
* K - ratio of Potassium content in soil
* temperature - temperature in degree Celsius
* humidity - relative humidity in %
* ph - ph value of the soil
* rainfall - rainfall in mm
* label(rice,potato…)

1. Find the statistical values
2. Visualize data using histograms
3. Build KNN Classifier Model and evaluate the metrics
4. Consider the Fertilizers recommendation.csv

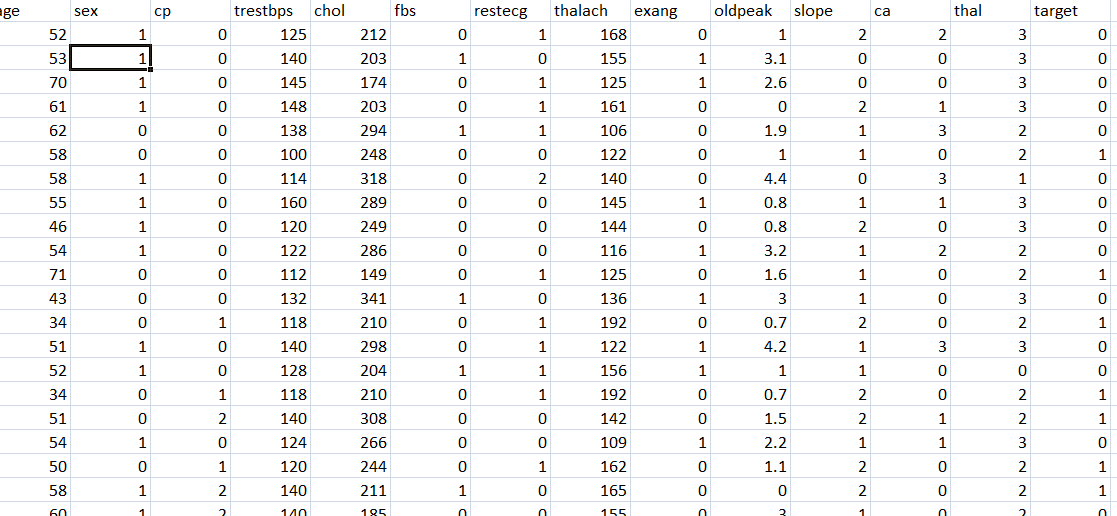


* Apply data Pre-processing techniques
* Visualize data
* Build the Random Forest classifier along with feature selection
* Analyze the performance metrics

1. Any Apply Data pre-processing techniques and build Ridge regression on any data set,visualize data using scatterplot

Evaluate the metrics

1. Consider the heart disease dataset
2. age
3. sex
4. chest pain type (4 values)
5. resting blood pressure
6. serum cholestoral in mg/dl
7. fasting blood sugar > 120 mg/dl
8. resting electrocardiographic results (values 0,1,2)
9. maximum heart rate achieved
10. exercise induced angina
11. oldpeak = ST depression induced by exercise relative to rest
12. the slope of the peak exercise ST segment
13. number of major vessels (0-3) colored by flourosopy
14. thal: 0 = normal; 1 = fixed defect; 2 = reversable defect  
    The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.



* Apply PCA feature selection (Write Steps)
* Build SVM classifier
* Analyze the performance metrics

1. Apply Pre-processing techniques

Apply Adaboost classifier on Diabetics

Visualize data

Find the performance metrics

1. Apply SVD and KNN on Diabetics

Write the Mathematical procedure of SVD and PCA algorithm

Find the performance metrics

1. Any Apply Data pre-processing techniques and build Polynomial regression on any data set

Evaluate the metrics

1. Consider the Crop Recommendation Dataset.csv

This dataset was build by augmenting datasets of rainfall, climate and fertilizer data available for India.

Data fields

* N - ratio of Nitrogen content in soil
* P - ratio of Phosphorous content in soil
* K - ratio of Potassium content in soil
* temperature - temperature in degree Celsius
* humidity - relative humidity in %
* ph - ph value of the soil
* rainfall - rainfall in mm
* label(rice,potato…)

1. Find the statistical values
2. Visualize data using boxplots
3. Build RandomForest Classifier Model and evaluate the metrics
4. Consider any Dataset.csv
5. Find the statistical values and Apply any wrapper based feature selection techniques
6. Visualize data
7. Build SVM Classifier Model and evaluate the metrics
8. Consider the Numerical Dataset.csv
9. Apply any filter based feature selection techniques
10. Find oultliers using bxplots and visualized
11. Build Ridge and multi linear regression Model and evaluate the metrics