Crop Yield Prediction

Open Data project - Report

Supervisor - Dr. Gagandeep Kaur, Ms. Sarishty Gupta

Enrollment No.	Name	Batch
15103297	Jitesh Pabla	B8
15103311	Vaibhav Sharma	B8
15103332	Sajal Subodh	B8

Introduction

The history of Agriculture in India dates back to Indus Valley Civilization Era and even before that in some parts of Southern India.India ranks second worldwide in farm outputs. Agriculture and allied sectors like forestry and fisheries accounted for 15.4% of the GDP (gross domestic product) in 2016 with about 31% of the workforce in 2014. India ranks first globally with highest net cropped area followed by US and China.The economic contribution of agriculture to India's GDP is steadily declining with the country's broad-based economic growth. Still, agriculture is demographically the broadest economic sector and plays a significant role in the overall socio-economic fabric of India.

The reason for this decline in the agriculture sector is due to the fact that farmers are not empowered and due to lack of application of IT in the farming sector. Farmers have less knowledge about the crops they grow.

We tend to overcome this obstacle by applying machine learning techniques to predict the crop yield and name by considering various factors such as temperature, rainfall,
season and area.

Application Domain and Goals

The project finds a huge application in improving real life farming scenarios. A lot of crop is destroyed every year due to lack of knowledge of weather patterns such as temperature, rainfall, etc. which play a huge role in deciding the crop yield. This project not only helps in predicting these parameters for throughout the year , but also assists in predicting the yield of various crops in various seasons based on past trends. Hence it allows the farmers to decide the best crop to grow to suffer minimum losses.

Data Source and Datasets

The acquiring of dataset in the Indian sub terrain is a tad difficult as there is no official compilation of the required datasets but scattered datasets are available which upon merging can be used to provide the desired yield.

The following data sets were used throughout the project:

• **Temperature and Rainfall** : These datasets were collected from Indian Water Portal (www.indianwaterportal.org). They consists of 100 years of temperature/rainfall data per month for each district in India as follows :

District	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
adilabad	1912	6.725	10.488	23.288	35.56	23.119	115.546	294.119	276.865	181.615	47.31	1.339	0
adilabad	1913	0.42	0	0.388	6.07	3.331	45.96	233.973	167.971	198.177	26.447	35.083	11.222
adilabad	1914	6.643	1.956	0.173	4.551	33.348	132.078	436.611	334.544	226.037	138.818	14.095	8.823
adilabad	1915	0.054	0.121	11.446	0.017	16.9	131.048	160.694	81.865	251.577	110.391	0.146	0.13
adilabad	1916	0.589	2.293	8.252	35.02	17.569	79.937	96.331	313.522	361.697	4.95	0.146	0
adilabad	1917	4.369	1.967	9.703	0.326	2.983	221.619	279.633	189.606	132.764	11.557	23.836	6.166
adilabad	1918	0.921	2.758	2.275	27.166	0.537	141.804	221.362	301.701	83.428	0.735	4.585	3.344
adilabad	1919	2.478	5.536	3.84	3.644	8.194	100.301	271.8	333.651	364.372	0.715	0.146	0.036
adilabad	1920	0.265	2.163	3.161	42.729	4.311	134.261	284.987	174.845	121.895	1.791	0.146	7.174

Rainfall Data

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	District
1912	22.965	24.17	28.024	31.609	33	31.1	27.18	25.999	27.661	26.686	23.167	21.096	Adilabad
1913	23.233	24.964	29.675	32.378	35.101	32.33	28.376	27.853	26.852	26.614	22.351	21.26	Adilabad
1914	22.708	24.058	28.554	32.16	32.434	32.286	27.008	26.479	26.762	25.25	21.657	20.216	Adilabad
1915	22.413	24.189	28.269	32.898	33.69	29.975	27.588	27.1	26.565	26.117	22.051	20.894	Adilabad
1916	21.963	22.996	28.156	30.432	34.193	32.5	27.96	26.899	26.491	26.089	23.558	20.526	Adilabad
1917	22.599	24.932	27.901	32.973	35.192	30.038	27.404	26.833	26.482	25.864	23.249	21.637	Adilabad
1918	22.171	24.722	27.891	29.792	32.966	30.733	27.861	25.406	27.432	27.157	24.343	20.63	Adilabad
1919	21.402	24.856	27.817	33.023	34.671	31.838	26.92	25.992	26.707	25.976	22.181	19.461	Adilabad
1920	23.312	25.207	29.182	29.945	33.747	29.65	26.531	27.075	26.535	26.098	23.249	22.181	Adilabad
1921	22.401	24.173	28.672	32.754	34.235	30.022	27.908	26.924	26.327	25.628	21.175	20.16	Adilabad

Temperature Data

- **Crop Yield Data :** This dataset is available on the Open Government Data (OGD) Platform of India (www.data.gov.in) .This consists of data points of 17 years that is categorised on the basis of districts, crops, seasons and area.

State_Name	District_Nam	Crop_Year	Season	Crop	Area	Production
Andaman an	NICOBARS	2000	Kharif	Arecanut	1254	2000
Andaman an	NICOBARS	2000	Kharif	Other Kharif	2	1
Andaman an	NICOBARS	2000	Kharif	Rice	102	321
Andaman an	NICOBARS	2000	Whole Year	Banana	176	641
Andaman an	NICOBARS	2000	Whole Year	Cashewnut	720	165
Andaman an	NICOBARS	2000	Whole Year	Coconut	18168	65100000
Andaman an	NICOBARS	2000	Whole Year	Dry ginger	36	100
Andaman an	NICOBARS	2000	Whole Year	Sugarcane	1	2
Andaman an	NICOBARS	2000	Whole Year	Sweet potatc	5	15
Andaman an	NICOBARS	2000	Whole Year	Tapioca	40	169
Andaman an	NICOBARS	2001	Kharif	Arecanut	1254	2061
Andaman an	NICOBARS	2001	Kharif	Other Kharif	2	1
Andaman an	NICOBARS	2001	Kharif	Rice	83	300
Andaman an	NICOBARS	2001	Whole Year	Cashewnut	719	192

- **The Final Dataset**: This data set is made by merging the temperature, rainfall and production into the final .csv file. The predictions were made on this final data set to get the required results.

State	District	Year	Season	Crop	Area	Production	emperatur	Rainfall
Andhra Pra	anantapur	1997	Kharif	Arhar/Tur	21400	2600	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Bajra	1400	500	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Castor seec	1000	100	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Cotton(lint)	7300	9400	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Dry chillies	3700	7100	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Groundnut	650800	228400	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Horse-gram	3300	1000	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Jowar	10100	10200	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Korra	2200	700	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Maize	2800	4900	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Moong(Gre	1300	500	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Other Khari	800	100	25.52133	68.57033
Andhra Pra	anantapur	1997	Kharif	Ragi	6700	11800	25.52133	68.57033

Algorithms Applied

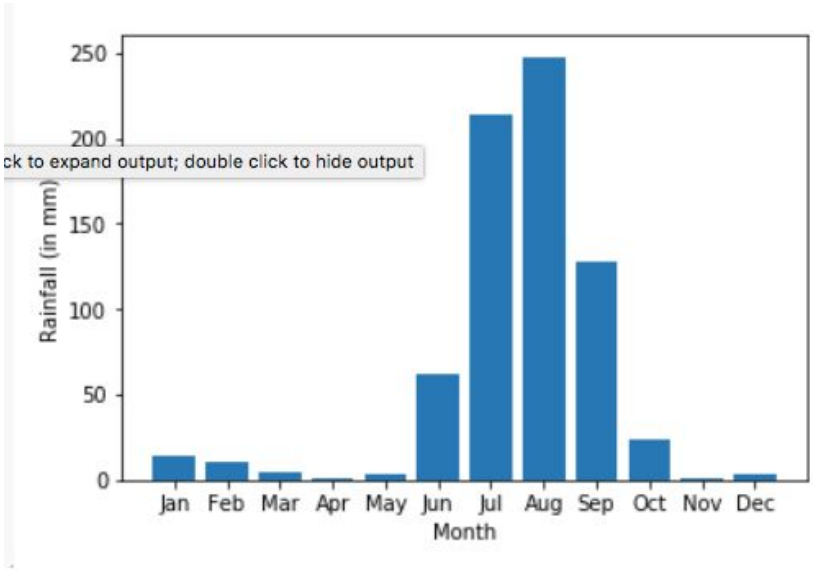
- **Rainfall and Temperature:**
  The data was sequential in nature thus time series analysis algorithms were used:
  1. **LSTM**: LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series.
  2. **Simple RNN**: A recurrent neural network (RNN) is a class of artificial neural network where connections between nodes form a directed graph along a sequence. This allows it to exhibit temporal dynamic behaviour for a time sequence.
  3. **Linear Regression**: Forecasting by minimizing the errors in prediction in past.

- **Final Dataset** :
  Regression Analysis and Classification Analysis were both applied on this data set.
  1. KNN Regressor and classification
  2. Artificial Neural Net
  3. Linear Regression
  4. SGD Regressor and classification
  5. Random Forest Regressor and classification

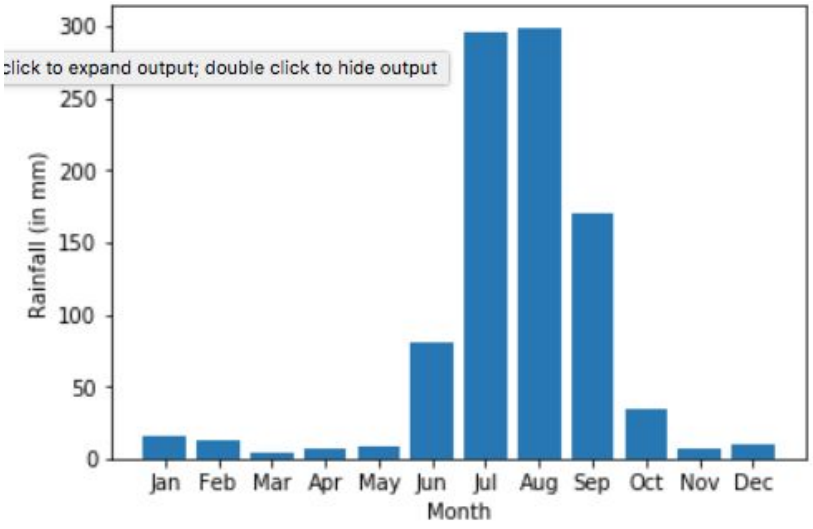
Comprehensive Analysis of the Algorithms Applied

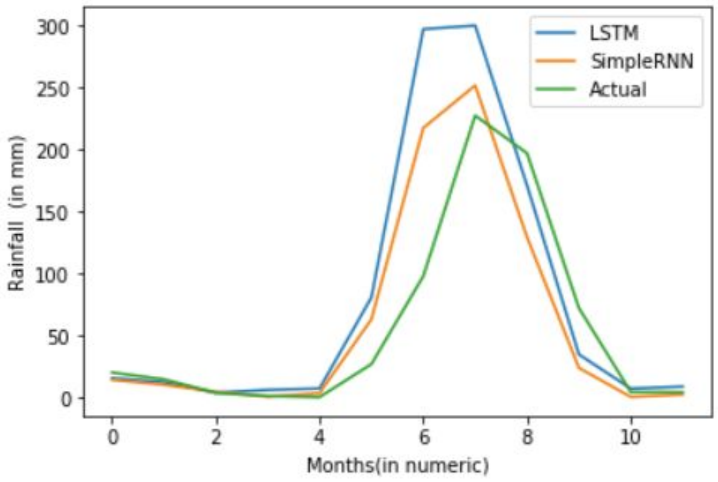
• **Rainfall Prediction:** We have taken a look back of 10 years and thus predicting the rainfall for 11th year for training data.

1. Simple RNN



2. LSTM





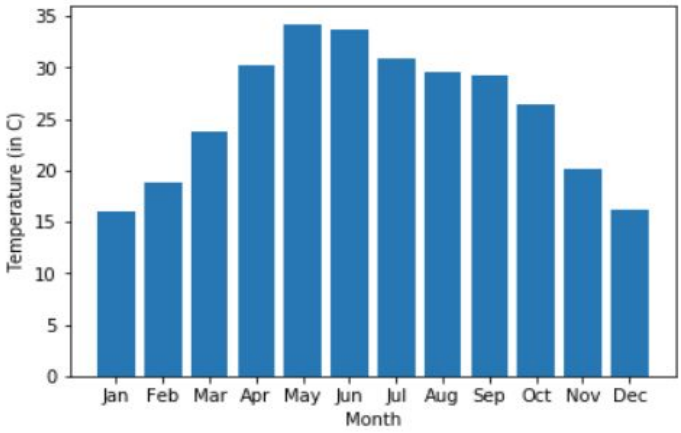
Comparison of actual distribution with predicted.

Model	Mean Absolute Error
Simple RNN	22.17
LSTM	34.14

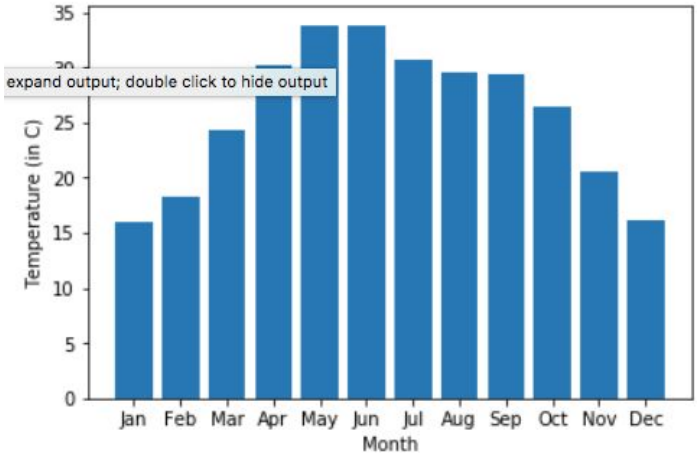
Comparison of mean absolute error for each model

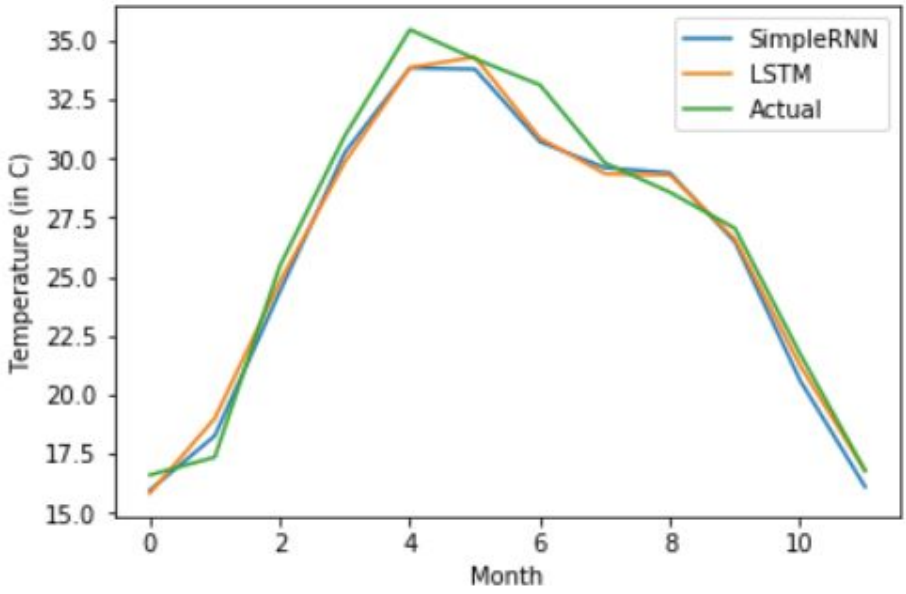
• **Temperature Prediction:** We have taken a look back of 10 years and thus predicting the temperature for 11th year for training data.

1. Simple RNN



2. LSTM





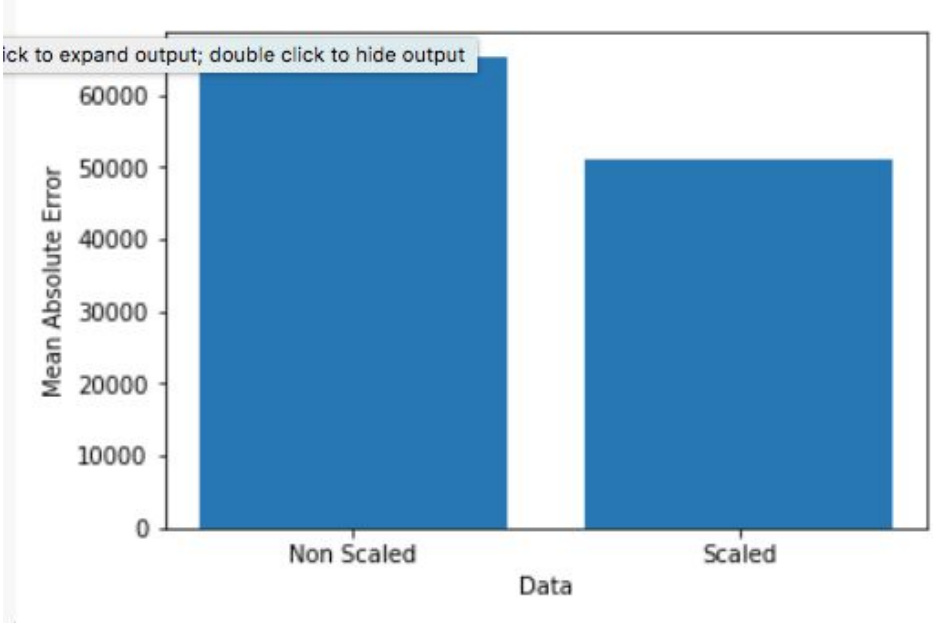
Comparison of actual distribution with predicted.

Model	Mean Absolute Error
Simple RNN	0.92016
LSTM	0.8137

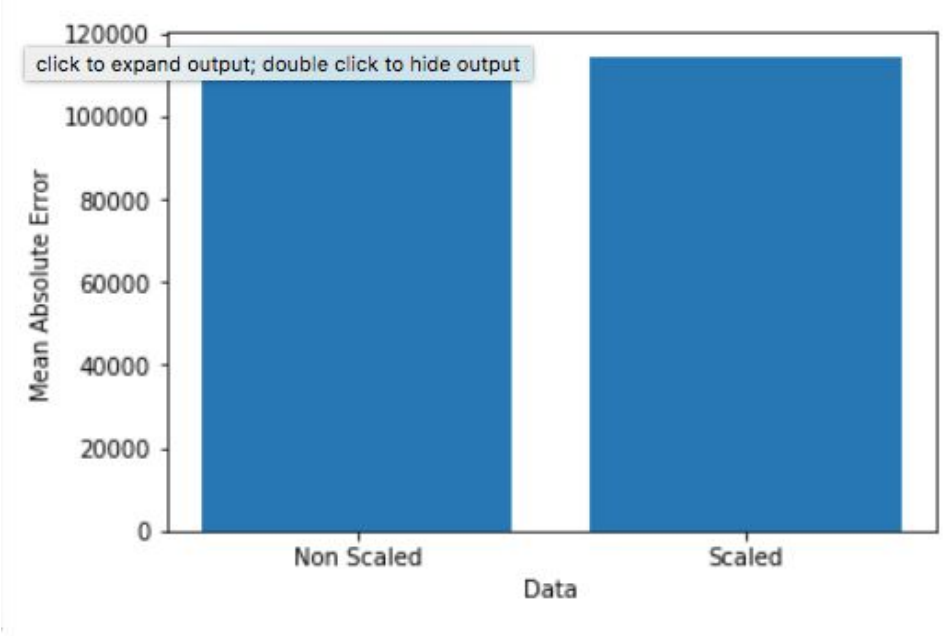
Comparison of mean absolute error for each model

• **Yield Prediction:** We are predicting the value of prediction and then dividing it by the area of the chosen district thus in turn predicting the crop name.

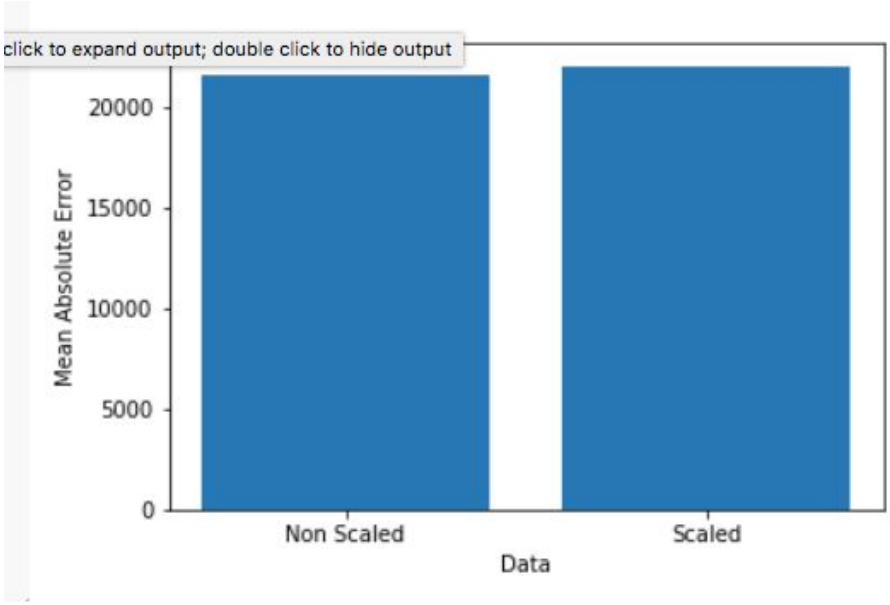
1. Artificial Neural Net



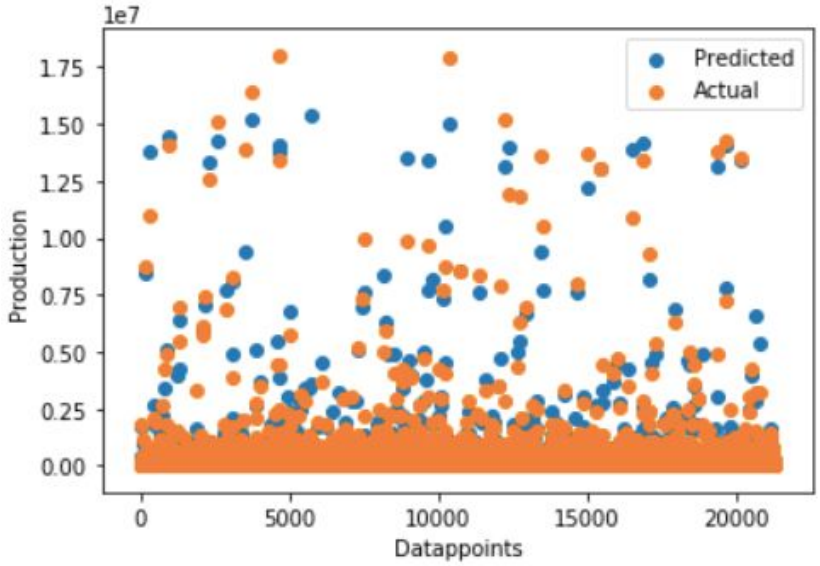
2. Linear Regression



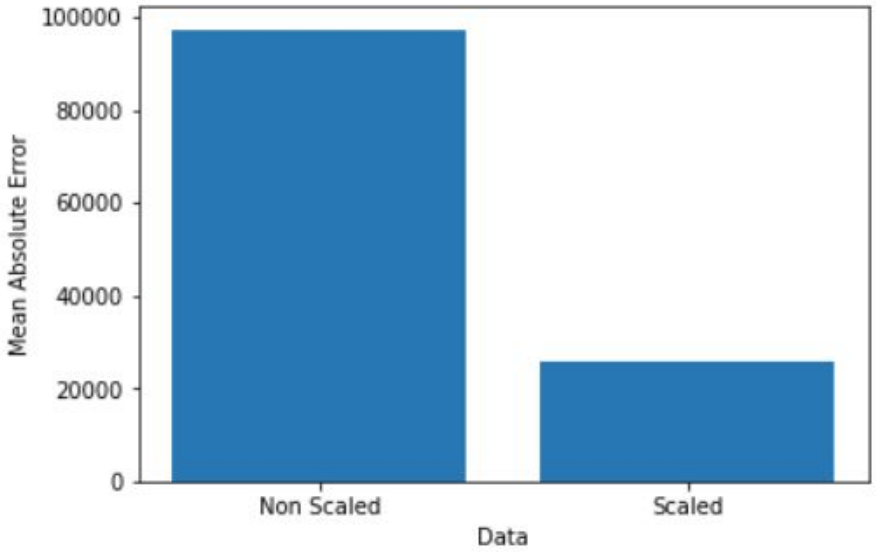
3. Random Forest Regressor



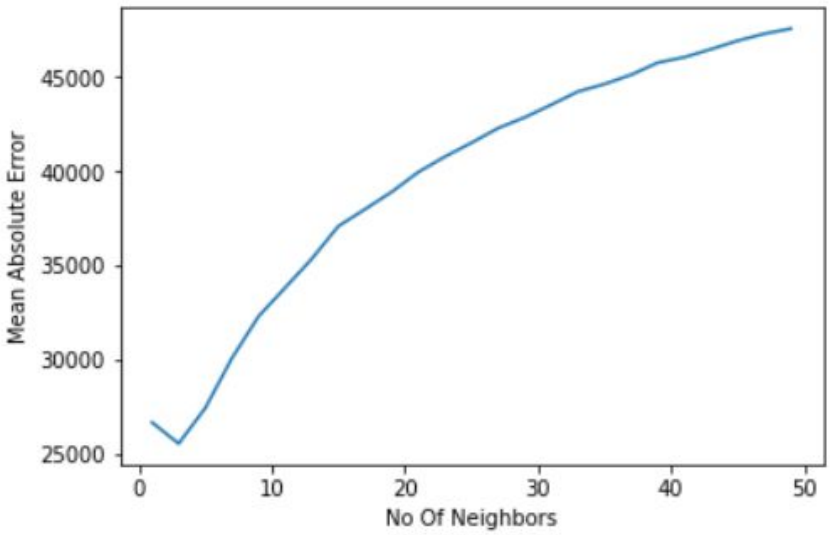
Predicted vs Actual Production Value for Random Forest Regressor



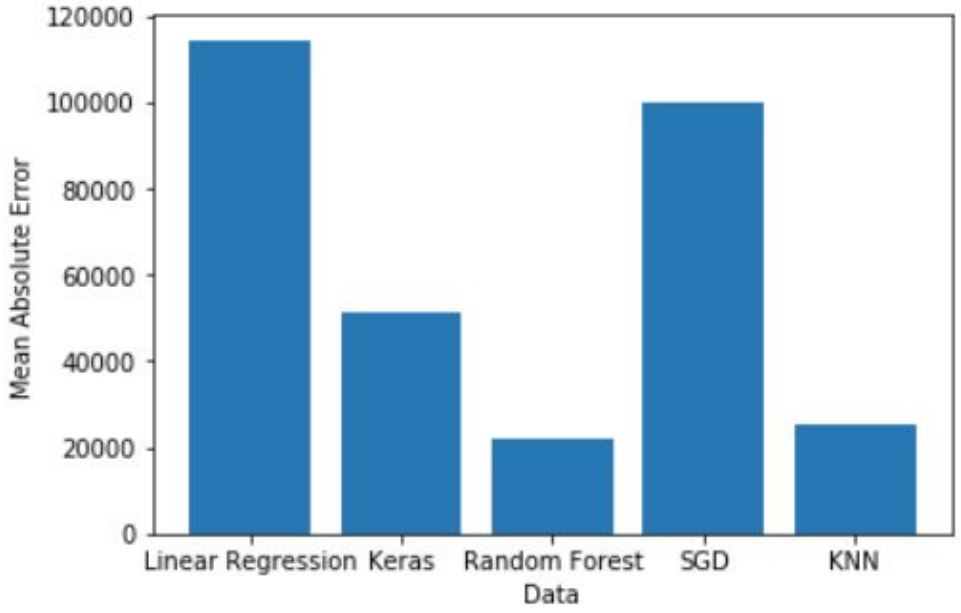
4 . KNN Regressor



Mean Absolute Error varying in accordance with number of neighbours (KNN)



Comparison of various techniques



Result Analysis

• The dataset size is around 1 lakh.
• The best technique for rainfall is SimpleRNN with a mean absolute error of 22.14mm
• The best technique for temperature is LSTM with a mean absolute error of 0.8137 degrees.
• After applying various techniques we found out that in Crop Yield annd Crop Name Random Forest yields the best result with minimum mean absolute error.

Conclusion and Future Scope

This project solves one of the fundamental problems that the Indian farmers are facing that is selection of which type of crop will yield the maximum results.The sole objective is to increase farmer's income.

Lack of proper dataset is the major hurdle while predicting the name of the crops but we were able to manage that by merging different data sets.

This project right now covers only five features that are season ,area ,temperature, rainfall and crop name but that's not the end, this project holds numerous possibilities such as the addition of vapour pressure, soil quality and market integration.

This project if compiled with a bigger data set can be a boon for the government as it may help them plan properly and in turn help our objective.

References

- https://data.gov.in

- http://www.indiawaterportal.org/questions/need-rainfall-and-temperature-data-state-and-district-wise

- Subhadra Mishra, Debahuti Mishra, Gour Hari Santra. Applications of Machine Learning Techniques in Agriculture Crop Production: A Review Paper, Oct 2016.

- P.Priya, U.Muthaiah, M.Balamurugan. Predicting yield of the crop using Machine Learning Algorithms.