PAPER • **OPEN ACCESS**

Indian agriculture GDP and non performing assets: A regression model

To cite this article: Bhoomika Jethwani *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1042** 012007

IOP Conf. Series: Materials Science and Engineering 1042 (2021) 012007 doi:10.1088/1757-899X/1042/1/012007

Indian agriculture GDP and non performing assets: A regression model

Bhoomika Jethwani[1], Devasya Dave[2], Tamanna Ali[3], Shraddha Phansalkar[4] and Swati Ahhirao[5]

[12345]Department of Computer Science and Engineering, Symbiosis Institute of Technology, Lavale, Tal Mulshi, Pune, 412115, India
bhoomika.jethwani@sitpune.edu.in, devasya.dave@sitpune.edu.in, ali.tamanna@sitpune.edu.in, shraddhap@sitpune.edu.in, swatia@sitpune.edu.in

Abstract. Agriculture is one of India's crucial sectors in terms of its contribution to employment and the country's (Gross Domestic Product) GDP. It has primarily emerged as an essential -growing sector in the global economy since independence. [21] However, the non-realization of the reasonable price for agricultural crop production leads to the introduction of loan waivers, which impact the credit culture and weaken the farming economy and growth. The presented work aims to perform exploratory data analytics on the GDP data in agriculture public domain by performing feature engineering on the factors affecting the agricultural GDP using the data for the period 1961 to 2019. It further builds a multi-linear prediction model to forecast the Agriculture Sector's economic performance in terms of GDP and NPAs generated by the Agricultural Sector using Machine Learning Techniques. Keywords: Multiple Linear Regression, Ordinary least Squared, Agricultural GDP, Non-Performing assets, Indian Economy.

1. Introduction

In today's day, the identification of nations of the world is made by their economy's strength. High wage jobs and improved quality of life are some components that are driven by economic development. We can portray the economy's growth using indicators that attempt to present an overall idea of any particular geography's financial health. Economic growth is an essential indicator because it indicates growth, whether measured by GDP (Gross Domestic Product), GVA (Gross Value Added).

GDP data helps to measure economic health and future inflation. High GDP data portrays the strength of an economy. Higher growth increase investors automatically increase the country's exports, which have a direct impact on the jobs and infrastructure. [6] However, a shrinking GDP is an early sign of recession, decreasing wages, and declining business revenues. Thus, the GDP information is crucial for the health of the economy and everything else that is remotely associated with it. A flourishing economy's skeleton is a strong banking sector, failure of which can cause adverse effects on other industries, as well. In India, banks face lots of concerns regarding India is Non-Performing Assets (NPA). The rise in NPAs implies a high possibility of increased credit defaults that, in turn, affect banks' profitability and net worth and plunge the asset's value. Thus, there is an indispensable need to control banks' issues regarding NPA's in the Agricultural Industry.

Banks must identify whether crediting loans into a particular sector is beneficial or not to avoid NPAs. This work intends to study the agricultural sector's performance in terms of GDP and bad loans (NPA).

IOP Conf. Series: Materials Science and Engineering 1042 (2021) 012007 doi:10.1088/1757-899X/1042/1/012007

Contributions of this work:

1. Recognizing key features affecting the agricultural GDP.
2. Analyzing the performance of the Agricultural sector in terms of GDP.
3. Proposing a model that predicts the percentage of NPA in terms of Agricultural sector GDP.

2. Related Work

The NPA is a short form of 'Non-Performing Asset.' The built-up of NPA has been the dominant factor in destroying the public sector banks' profitability in India. [2] Narasimham Committee (II) has emphasized the need to reduce all banks' average NPAs. It has recommended prudential norms on income recognition, asset classification & provisioning. When interest or other dues to a bank remains outstanding for longer than 90 days, the NPA rule affirms that the entire bank loan automatically turns into a non-performing asset.

The major problem for banks and financial institutions is the recovery of loans. The plan of action is firstly avoiding NPA, and if unavoidable, then identify and manage the factors responsible for it.

[2] Emulating international best practices and ensuring greater transparency, the 90 days' overdue criterion for recognizing NPAs has been selected, from the fiscal year 2005. Consequently, a non-performing asset (NPA) will be a loan or an advance where:

i. The principal's interest and installment remain overdue for more than 90 days (for term loan).
ii. Said account remains out of order for more than 90 days, in respect of an Overdraft/Cash Credit.
iii. The bill remains unpaid and overdue for more than 90 days in the case of purchased and discounted bills.
iv. Interest and installment of the principal continue to be unpaid and overdue for two harvest seasons.

The problem of Non Performing Assets has been a matter of discussion for commercial systems worldwide. [28] NPA's harm the banks and, consequently, the whole economy. The state of well-being of the production and trade is reflected by the level of NPA's in Indian banks. The prime duty of banking is the granting of credit. However, lending also carries a risk called credit risk, which arises from the borrower's failure. The non-recovery of loans causes a significant hurdle in the process of the credit cycle. These loans affect the bank's profitability. The problem of Non Performing Assets has been a matter of discussion for commercial systems worldwide. [28] NPA's harm the banks and, consequently, the whole economy. The state of well-being of the production and trade is reflected by the level of NPA's in Indian banks. The prime duty of banking is the granting of credit. However, lending also carries a risk called credit risk, which arises from the borrower's failure. The non-recovery of loans causes a significant hurdle in the process of the credit cycle. These loans affect the bank's profitability. NPA control is the most significant hurdle encountered by the Banking Industry in India. Non Performing Assets (NPA) are a loss to the economy. Although the causes of NPA are varied, one of them is poor credit decisions by bank management. These decisions are taken due to the unavailability of data, inability to forecast the growth in a particular sector, policies of direct credit, that is, the crude form of behest lending.

In Sergio's work, [24] in a study on non-performing loans in Italy, an increase in loan assets' riskiness is rooted in a bank's lending policy adducing relative unselective and inadequate evaluation of sectorial prospects.

In Gopalakrishnan's work, [26] the causes of NPAs can be classified into political, economic, social, and technological factors. He perceived that oversight of regular credit appraisal, absence of follow-up and surveillance, recessional stresses in the marketplace, a shift in government policies, and digression of funds are the major causes of NPAs.

IOP Conf. Series: Materials Science and Engineering 1042 (2021) 012007 doi:10.1088/1757-899X/1042/1/012007

[23] Challenges faced by the agricultural sector:

1. Fragmentation and uneconomical size of land holdings due to the *Zamindari* system and division of land amongst family members through numerous generations.
2. High wastage due to poor storage and supply chain infrastructure.
3. Significant dependency on monsoon and insufficiency of irrigation facilities.
4. Lack of extension duties in the sector, which include applying scientific research and new knowledge to agricultural practices.
5. Governments debt waiver schemes
6. Stress in the sector is inadequate to justify the growth in non-performing loans. According to the report of Standard Chartered Securities, "*The problem with farm loans is not slow growth but rising non-performing loans as farmers expect more debt waive*".
7. Mandatory requirements by the government
8. An increase in NPA can be attributed mainly to existing policy prescriptions, substantially increasing agricultural credit.

However, no systems are analyzing the primary reasons contributing to high NPA ratios in agricultural finance and how the non-repayment of loans affects performance and the GDP.

3. Proposed Work

We propose a model that finds the key features relating to GPD and NPA, which can be tackled to solve the problem of NPA management in India. In this work, we focus on the agricultural sector, and the system is based on the following objectives:

i. Prediction of Agricultural GDP using the regression model.
ii. Prediction of NPA in the agricultural sector.
iii. Analyzing the performance of Agricultural GDP for the period 1961 - 2019.
iv. Investigating the States contributing to the maximum amount of NPA and the types of loans provided to farmers that often tend to go wrong.

3.1 Methodology

This work explores the capabilities of ML methods in the task of predicting the growth and presence of bad loans in the agricultural sector. We propose the method of supervised ML Regression algorithms in a two-phase process for this purpose. In the first phase, the dataset is split into the training and testing data. The second phase involves the application of the regression algorithms to build and identify the best model. This permits us to recognize the most relevant results and thereby increase accuracy.

3.1.1 Dataset Construction. The first phase of data analysis for predicting growth and presence of bad loans in the agricultural sector is collecting the data on essential features from various open sources, dataset repositories, official annual reports, and research papers. However, the collection and integration of data from such varied sources are challenging. A labeled dataset is required to develop an efficient prediction model. We have prepared such a dataset by manually labeling the features that will enable us to predict the relevant output.

Variables used in the dataset are as follows:

Area under agriculture (million hectares)
India GDP (billion USD)
Export Value (billion USD)
LokSabha Elections (Yes/No)
Outstanding credit amount (billion USD)
Agriculture Percentage in India GDP
Arable Land (million hectares)
Agriculture GDP (billion USD)
Fertiliser Consumption (kg/hectare)
Area under agriculture (million hectares)
Rainfall (mm)
Year (1960 - 2019)
Agricultural machinery (per 100 sq. km. of arable land)
Outstanding Credit as percentage of Agricultural GDP
Agriculture Credit Disbursed as amount of Agriculture GDP
Agriculture Credit Disbursed as percentage of Agriculture GDP

Total FDI amount (billion USD)
Percentage of Agricultural NPAs (Public Sector Banks)
Percentage of Agricultural NPAs (Private Sector Banks)
Rainfall deviation in Percentage from LPA

3.1.2 Exploratory Data Analysis. In this work, the raw dataset is transformed into standardized forms to develop efficient models. This standardization has taken place in multiple steps:

- ***Handling categorical data:*** The dataset consisted of only one categorical feature, namely 'LokSabha Elections'. To handle this feature, *LabelEncoder* has been used, where each column is converted to a number.
- ***Correlation amongst features:*** It helps in feature selection by showcasing the relationship between the features. Heat-maps have been used as they are a good visualization of the correlation. Observations have led to the inference that the features with high correlation with agricultural GDP are -Production, Yield, Export Value, Credit disbursed amount, Outstanding credit amount.
- ***Handling missing values:*** The features where the number of missing values were much higher than available values, such as '*Fertiliser Consumption*' and '*FDI Amount*', were dropped. For features with a few missing values, such as '*Machinery*', the data was filled using the mean of all available values.
- ***Feature Scaling:*** MinMax Scaler has been used to scale the features to normalise the data within a particular range and to speed up the calculations in the algorithm.
- ***Visualisation:*** The visualisation of data has led to the following inferences:
    1. Production, Yield, Export Value, Credit Disbursed, Outstanding Credit have a strong positive correlation with Agriculture GDP.
    2. Rural Population, Employment, Arable Land have a strong negative correlation with Agriculture GDP.
    3. All the variables that have high p-value (> 0.05), like Area, Rainfall, etc have very low correlation with Agriculture GDP, so they will be insignificant in our model building process.

3.1.3 Feature Selection. The first step in regression analysis is to determine the criterion variable. Following our objective and system, for our first model, the dependent variable is Agricultural GDP. For our second model, the dependent variable is Outstanding Credit as a percentage of Agricultural GDP. The independent variables for each of the model needs to be chosen after the dependent variable is found. For an efficient model, it is advised to have the minimum number of features that cause maximum variance in the dependent variable.

This work has incorporated multiple feature selection components such as RFE, VIF, and p-value.

- Recursive Feature Elimination (RFE): It removes attributes recursively and builds a model on the remaining attributes. [25] An accuracy metric is used to rank the features according to their importance. The model to be used and the number of required features are given as input to the RFE method. Then, the ranking of each variable (1st feature is most important) is returned. It also provides its support where 'True' is relevant, and 'False' is irrelevant.
- Variance Inflation Factor (VIF): The collinearity of two variables can be identified using this. A VIF of 1 indicates non-collinearity, i.e., no correlation. However, the higher the VIF, the higher is the collinearity.
- P-value: [19] A low p-value signifies a more significant effect on the output, and the result is of greater importance. Statistically, having a p-value < 0.05 is deemed vital. In this work, we have selected the features with a p-value of less than 0.05 for each model.

For both of the models in our system, feature selection has been done in two steps.

A) *Agricultural GDP Model*

i. Features with insignificant p-value such as 'percentage of NPAs in Public Sector Banks and Private Sector Banks', 'Arable land', 'Rainfall Deviation' were removed.

ii. Feature with non-coherent VIF such as 'Yield' has been removed.

B) *NPA Model*

i. Features with insignificant p-value such as 'Rural Population', 'Export Value' were removed.

b. Feature with non-coherent VIF such as 'Outstanding Credit Amount' has been removed.

4. Trends of Agricultural GDP

Depicting the correlation of features with the Agricultural GDP and analyzing the p-value helps us select features to build an accurate prediction model. The graphs of features having a linear relationship with GDP and p-value<0.05 are further analyzed along the period (1960 - 2019) below.

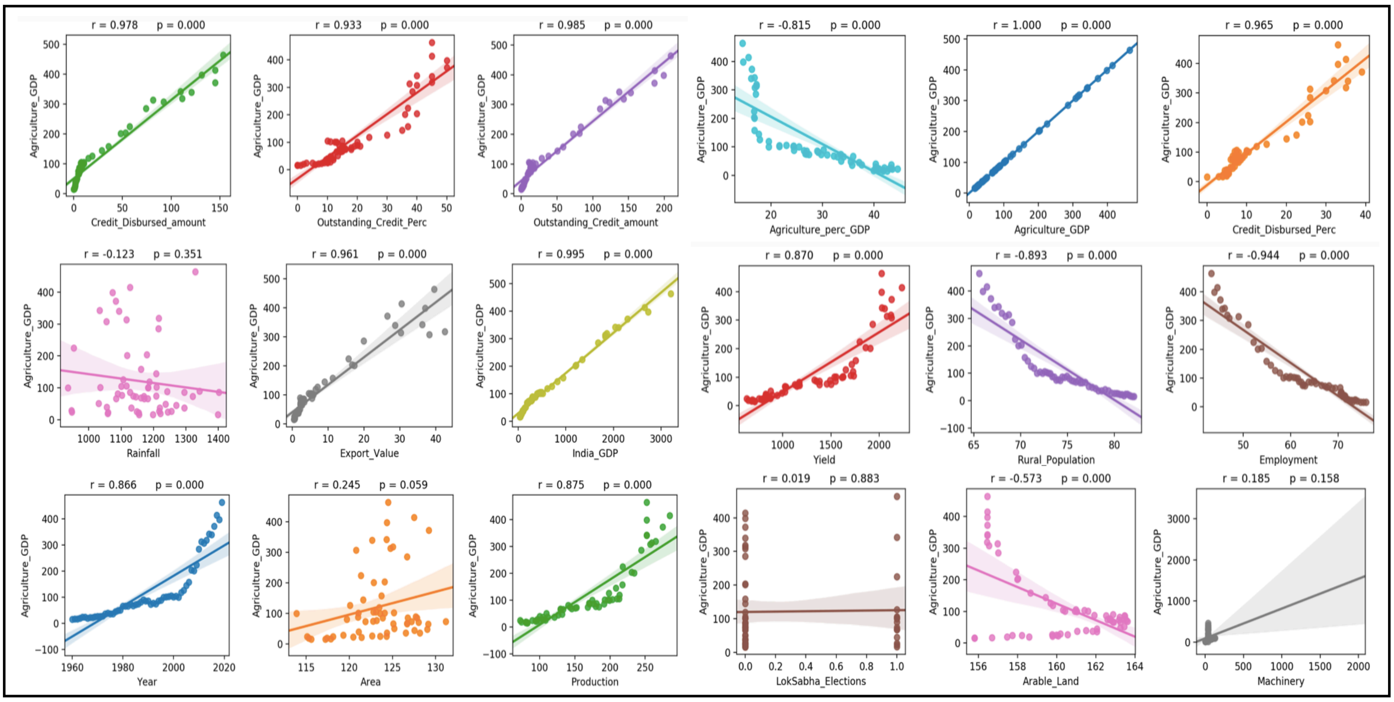

Figure 1: Correlation of features with Agricultural GDP

4.1 Credit disbursed vs. outstanding credit as a percentage of Agricultural GDP

The ratio of Agricultural Credit outstanding to Agricultural GDP peaked from 0.6 % in 1950-51 to 9.81 % in 1971-72. [7]The rate shows an escalated trend increasing to 21.76 %, from 1972 to 1986-87. The remarkable success of agricultural credit vs. agricultural GDP from the 1950s-1980s is based on banks' nationalization and the initiation of RRBs. However, the backward trend in the ratio commenced from 1991 onwards, and it fell to 13.34 % in 1998. Following 1999, the rate heightened steeply and reached up to 39.55 % in 2006-07, which indicates that the entrance of KCC was a big promoter for agricultural credit and induced a significant transformation in widening the reach of credit to the farming population. The loan waivers' announcement negatively influenced the borrowers' reparation behavior and made the banks reluctant to fresh lending.

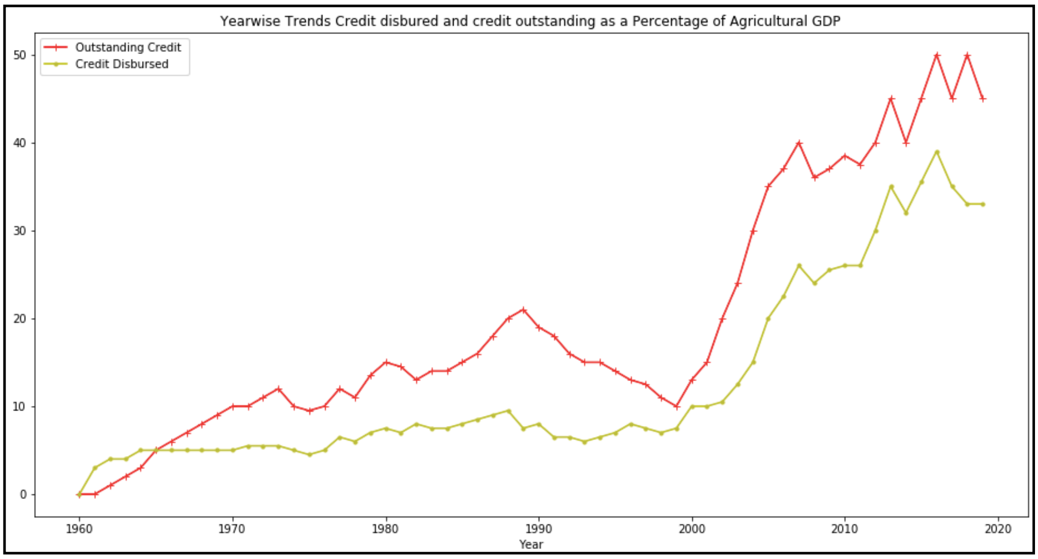

Figure 2: Credit disbursed vs. outstanding credit year wise trends

4.2 Export Value of crops

India is second in the global farming generation, but its share of world agricultural exportation is as low as 2%. Agricultural exports' benefaction to India's GDP is also as little as 2%, weaker than other developing agrarian country. Since the economy reopened in 1991, [27] India's agricultural business surplus listed more than ten-fold rise between 1991-92 and 2013-14. The brisk pace at which exports increased to balance the corresponding increase in imports. However, in the last few years, agricultural exports fell by 22% while imports increased by 62%, resulting in a significant drop in a trade surplus of 70%, consequently affecting the agricultural GDP.

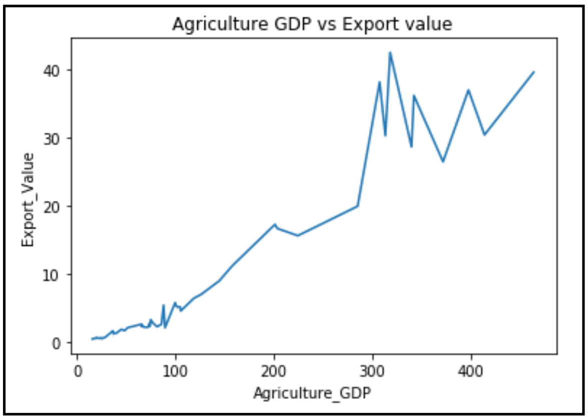

Figure 3: Export Value Trends vs Agri-GDP

4.3 Rural Population of India

[12.] In 2018, the rural population in India was 892,321,700. Over the past 58 years, it reached a maximum of 892,321,700 in 2018 and a minimum amount of 369,791,500 in 1960. In rural areas, agriculture plays an essential role in the economy. [21] Over 70 percent of rural families depend on agriculture for their daily sustenance. It employs over 60% of the rural population. Indian agriculture has registered remarkable growth over the last few decades, along with its inexorable rise in population. Hence, it is clear that the growth of the population is linearly related to the Agricultural GDP.

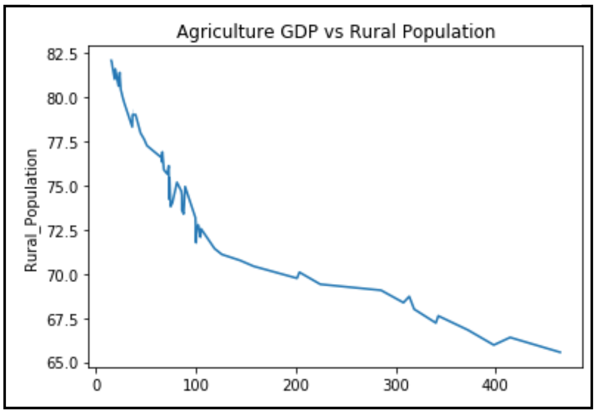

Figure 4: Population Trends vs Agri-GDP

5. Trends of Agricultural NPA

Loan enforcement of PSBs shows co-movement with rainfall deficiency, except in 2008-09 and 2009-10, where loan write-offs could have influenced the lessening of the Non- Performing Assets (NPA) level despite high rainfall scarcity. [7] In the recent years of 2016-17 and 2017-18, the NPA level has dramatically risen, possibly symbolizing strategic default stemming from the state-level loan waiver releases. The linkage between agrarian distress and unfavorable rainfall in India is well established, as the foundation for irrigation is still lacking. At the state level, rainfall performance in the years was typical, with no notable deviation from Long Period Average (LPA) between 2004 to 2008, after three straight years of drought between 2001 and 2003, causing a drop in GDP.

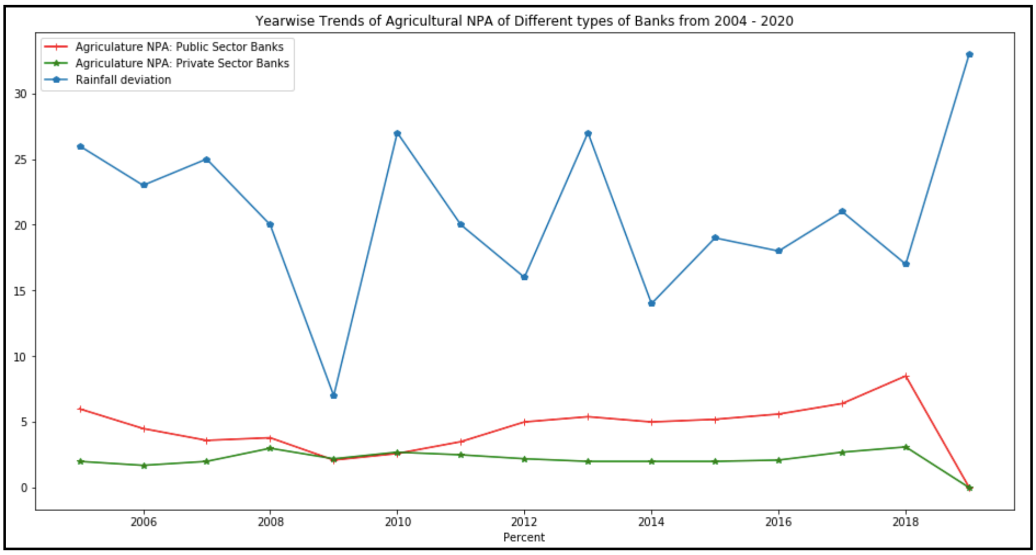

Figure 5: . Public sector NPA vs. Private sector NPA

5.1 State-wise data analysis

The data shows that the NPA level advanced for all states that have announced a farm loan waiver program in 2017-19. On the other hand, [7] almost all other states (except for Bihar, Odisha, and Haryana) have shown no significant change in their NPA level or have registered for a decline between 2016-18. Taken collectively, this could be suggestive of moral risk, with borrowers defaulting strategically in expectation of loan waiver. The states giving to the maximum amount of NPAs are Maharashtra (15.5%), Karnataka (10.3%), and Uttar Pradesh (10.2%)).

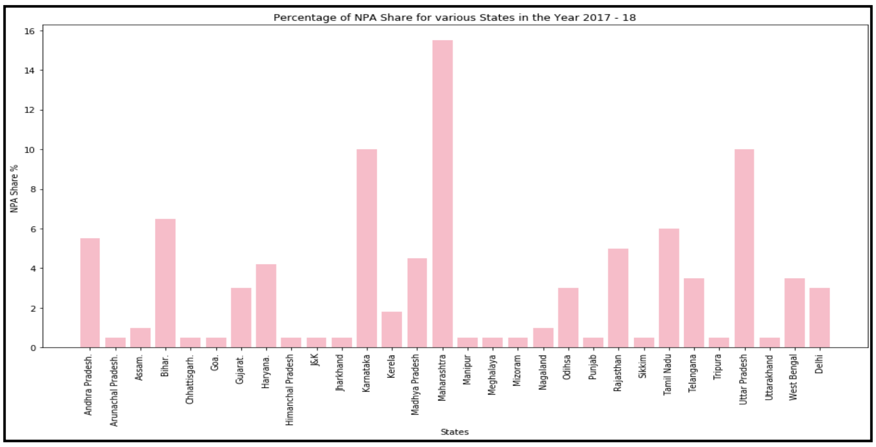

Figure 6: State-wise NPA

IOP Conf. Series: Materials Science and Engineering 1042 (2021) 012007 doi:10.1088/1757-899X/1042/1/012007

6. Proposed Model

Post data analysis, and feature selection, the dataset was fit into a Multiple Linear Regression algorithm to obtain the two prediction models. Since multiple independent variables are affecting the dependent variable, this model is used. The equation for multiple linear regression algorithm used in the model is:

$$y = b_0 + b_1 * x_{i1} + b_2 * x_{i2} + \ldots + b_n * x_{in} + e$$

In the above equation y is the dependent variable (here Agricultural GDP and NPA). Independent variables are x_{i1}, x_{i2} … x_{in} having coefficients/weights b_1 … b_n, b_0 is the y-intercept and e is the model's error term. OLS method is used to obtain a line that is closest to the actual data points. The intent of regression is to determine the value of coefficients, such that the error is insignificant. OLS strategy helps us to achieve that. The prediction model is designed after reducing its multi-collinearity and selecting features having a maximum correlation to the dependent variable.

7. Results and Discussion

After analysis, we can state that the key features affecting Agricultural GDP from the model (based on p-value) are the Rural Population of India, Credit amount Disbursed as Agricultural Loans, Export Value of Agricultural Goods.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        Agriculture_GDP   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.998
Method:                 Least Squares   F-statistic:                     1516.
Date:                Sun, 17 May 2020   Prob (F-statistic):           2.59e-34
Time:                        19:38:28   Log-Likelihood:                 65.611
No. Observations:                  42   AIC:                            -99.22
Df Residuals:                      26   BIC:                            -71.42
Df Model:                          15
Covariance Type:            nonrobust
==============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
const                         0.2430      0.063      3.845      0.001       0.113       0.373
Year                         -0.8429      0.209     -4.028      0.000      -1.273      -0.413
Area                         -0.0111      0.022     -0.498      0.622      -0.057       0.035
Production                    0.2178      0.070      3.116      0.004       0.074       0.362
Rural_Population             -0.9147      0.186     -4.922      0.000      -1.297      -0.533
Employment                    0.0537      0.151      0.355      0.725      -0.257       0.364
Rainfall                     -0.0222      0.012     -1.833      0.078      -0.047       0.003
Export_Value                  0.0694      0.027      2.576      0.016       0.014       0.125
India_GDP                     0.9596      0.114      8.383      0.000       0.724       1.195
Agriculture_perc_GDP          0.0943      0.051      1.862      0.074      -0.010       0.198
Credit_Disbursed_Perc        -0.0387      0.073     -0.530      0.601      -0.189       0.111
Credit_Disbursed_amount       0.3620      0.142      2.556      0.017       0.071       0.653
Outstanding_Credit_Perc       0.1627      0.072      2.256      0.033       0.014       0.311
Outstanding_Credit_amount    -0.6192      0.194     -3.192      0.004      -1.018      -0.220
LokSabha_Elections           -0.0024      0.004     -0.556      0.583      -0.011       0.006
Machinery                     0.0161      0.015      1.056      0.301      -0.015       0.048
==============================================================================
Omnibus:                       11.975   Durbin-Watson:                   1.455
Prob(Omnibus):                  0.003   Jarque-Bera (JB):               13.893
Skew:                           0.927   Prob(JB):                     0.000962
Kurtosis:                       5.122   Cond. No.                         199.
==============================================================================
```

Figure 7: Results for Agricultural GDP Model

Similarly, key features affecting Agricultural NPA from the model (based on p-value) are Area of agricultural land in India, Production of Agricultural crops, Employment rates and Credit Amount Disbursed as Agricultural Loans.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     Outstanding_Credit_Perc   R-squared:                       0.980
Model:                                 OLS   Adj. R-squared:                  0.972
Method:                      Least Squares   F-statistic:                     131.8
Date:                     Sun, 17 May 2020   Prob (F-statistic):           2.92e-22
Time:                             19:38:30   Log-Likelihood:                2.8892
No. Observations:                       42   AIC:                             18.22
Df Residuals:                           30   BIC:                             39.07
Df Model:                               11
Covariance Type:                 nonrobust
==============================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
const                      -0.3241      0.229     -1.416      0.167      -0.792       0.143
Year                        0.1704      0.391      0.436      0.666      -0.629       0.969
Area                        0.2702      0.076      3.555      0.001       0.115       0.426
Production                 -0.7758      0.238     -3.263      0.003      -1.261      -0.290
Employment                 -1.3162      0.541     -2.431      0.021      -2.422      -0.210
Rainfall                    0.0501      0.046      1.085      0.287      -0.044       0.144
India_GDP                  -0.0008      0.290     -0.003      0.998      -0.592       0.591
Agriculture_perc_GDP        0.0371      0.179      0.208      0.837      -0.328       0.402
Credit_Disbursed_Perc       0.8179      0.187      4.374      0.000       0.436       1.200
Credit_Disbursed_amount    -0.5982      0.288     -2.075      0.047      -1.187      -0.009
LokSabha_Elections         -0.0039      0.017     -0.228      0.821      -0.039       0.031
Machinery                  -0.1274      0.057     -2.238      0.033      -0.244      -0.011
==============================================================================
Omnibus:                        4.849   Durbin-Watson:                   1.981
Prob(Omnibus):                  0.089   Jarque-Bera (JB):                3.628
Skew:                           0.673   Prob(JB):                        0.163
Kurtosis:                       3.512   Cond. No.                         84.2
==============================================================================
```

Figure 8: Results for Agricultural NPA Model

7.1 Residual Analysis

Residual is computed by calculating the difference between the predicted output from the regression model and the measured output from the validation dataset. They represent the part of the validation data not defined by the model. Hence, we perform a residual analysis to check the accuracy of the model.

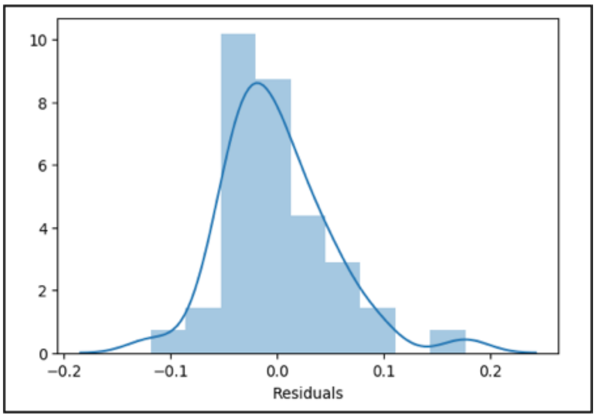

Figure 9: Residual Analysis for Agricultural NPA

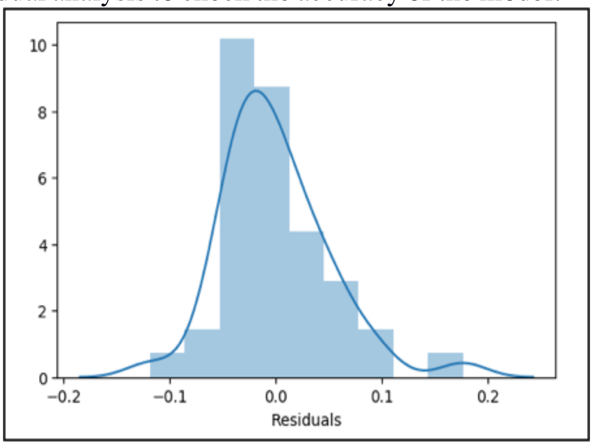

Figure 10: Residual Analysis for Agricultural GDP Model

As we can see, the residuals are normally distributed and have a mean of zero, which means that both the models are suitable.

7.2 Model Evaluation

Evaluation metrics used for the model are R Score and Mean square error to analyze its performance.

7.2.1 R score. The data points' closeness to the fitted regression line gives the R-squared statistical measure that ranges from (0 - 100 %). This value is also known as the coefficient of multiple determination for multiple regression. R-squared is f the percentage of the response variable variation that is explained by a linear model. A higher value of R-squared implies a better-fitted model for the data.

R-squared = Explained variation / Total variation

7.2.2 MSE (Mean squared error). MSE analyses how nearby a regression line is to the collection of data points. It calculates this value by taking the distances from the points to the regression line (these distances are known as "errors") and squaring them. Squaring the values helps to remove any negative signs. It also gives more weight to more significant differences.

Model	Accuracy (r score)	Mean Squared Error
Predict Agricultural GDP	0.995	0.011
Predict presence of bad loans	0.958	0.137

Table 1: Metrics for model evaluation

8. Application

The The intelligence from these models can be used to identify the growth and security of any sector. It can also be used to determine the number of threats any asset poses in terms of NPA. For example, should there be some agricultural sector assets, prone to turning into bad loans, then the necessary actions to manage them can be taken. The same holds for statistically predicting the presence and effect of these bad loans in the sector's GDP. Furthermore, should the data be analyzed in real-time, then the extracted intelligence will also be timely.

- The model notebook created can be further extended into an API for data-fetching applications or research extension purposes.
- The work can be deployed into a product with an interactive frontend to input and extract data resulting in an ever-growing dataset, to monitor the agricultural industry's performance.
- The concepts mentioned above can be released as an open-source repository for presenting contribution to data scientists worldwide to improve the accuracy and enhace the vision achieved in the current model.

8.1 Limitations of work

- The requirement to assemble a training dataset is achieved through manual labeling of the posts, which is resource-demanding, especially in the case of the multi-class dataset.
- Finally, the intricate features that have specific effects on the NPAs and the growth of a sector, as a whole, were not studied data-intensively.
- The model created cannot be reused as a starting point for creating models for other industries but only as a mark of citation due to similarity in use cases.
- The program does not provide a functionality to change the model algorithm as the nature and availability of data will change over the years to come.
- The Internet's available data is relatively inadequate since there is no proper record of data before 1960.

9. Conclusion

Agriculture is definitively one of the most influential contributors to our economy. [17] Despite such an overwhelming majority of the population engaged in agricultural activities, it is deplorable that the agrarian credit system is not yet well developed in India, and the existing systems are plagued with increasing numbers of NPAs. It is essential to realize that this is a systemic quandary that would not be fully solved by superfluous means such as farm loan waivers. It has to be tackled by resorting to some modern techniques and sophisticated mechanisms. These routines will help lessen the number of NPAs and further aid in boosting production levels, diminishing risks, and, consequently, leads to the overall growth of farmers' sustainable way. This work demonstrates the use of Multiple linear regression algorithms to predict economic growth in terms of GDP and bad loans as a percentage of GDP in the agricultural sector. [15] Even though most data analysts and financial practitioners have information about the NPAs and jeopardies involved in the industry's assets, distinguishing such assets and managing them has always posed a difficulty. Some features have been highlighted in this paper that affects the repayment of farm loans adversely like the rural population, the export value of crops, and crop production for that year. This consequently affects the GDP of the country. It also shows how identifying the features associated with a sector can help us determine its future growth and the risk involved in investing assets of those sectors. The acquired knowledge value also shows that this research also has a commercial development potential in the financial technology market that is yet unavailable.

10. Future Scope

This research work can provide a foundation for a more general framework for real-time data analysis from various sectors of the nation besides the agricultural industry. An interactive method that allows better visualization and exploration of the data would enable the analyst to understand better the risks and threats associated with assets. The relevance of this topic for future research lies in examining the extent of the adaptation of suggested technologies and mechanisms and further examining the impact of such adaptations on the NPA levels.

References

[1] Department of Agriculture, Cooperation and Farmers Welfare Annual Report 2018-29. http://agricoop.nic.in/sites/default/files/AR_2018-19_Final_for_Print.pdf
[2] Annual Report of the RBI 2018-19, Reserve Bank of India. https://www.rbi.org.in/Scripts/AnnualReportMainDisplay.aspx
[3] Economic Outlook. https://economicoutlook.cmie.com
[4] India, Knoema. https://knoema.com/atlas/India/topics/Agriculture
[5] Indiastat. https://www.indiastat.com/economy-data/8/union-budget/24469/2018-2019/1118397/stats.aspx
[6] Jayshree S, Focus on banking sector to tackle slowdown, Observer Research Foundation, 22 January, 2020. https://www.orfonline.org/research/focus-on-banking-sector-to-tackle-slowdown-60550/
[7] Reserve Bank of India’s “Report of the Internal Working Group to Review Agricultural Credit”
[8] Jha, Vijay Shekhar, “Problems of NPA in Banking Sector in India & Debt Recovery Remedies”, (May 1, 2018)
[9] Gopal, Sumathi, “NPA’s-A Comparative Analysis on Banks & Financial Institutions and its Implications- (Doctoral Thesis)”, (June 2010)
[10] S.K. Mishra, “Trends in growth of agriculture sector of the Indian economy”, (2007)
[11] Sagarika Rastogi, “Developing Block Chain Smart Contract Platform for the Sale of Stressed Assets in Banks along with NPA Analysis for various sectors through Predictive Models and Visualization Techniques”, (June 2019)

[12] India - Rural population, Index Mundi. https://www.indexmundi.com/facts/india/rural-population
[13] Pacha Malyadri, S. Sirisha, “Comparative Study of Non-Performing Assets in Indian Banking Industry”
[14] Kaveri.V.S., “Prevention of NPA suggested Strategies,” Vinimaya, 23(8), (2001)
[15] Prashanth. K. Reddy. “A comparative study of non-performing Assets India in the global context–similarities and dissimilarities, remedial measures,” (2002)
[16] JinWang, Zhiliang Chen, Kaiwei Sun, Hang Li, Xin Deng, “Multiple Target Regression via target specific features”, (2019)
[17] Ahita P, Examining the rise of Non-Performing Assets in India, PRS Legislative Research 13 September 2018. https://www.prsindia.org/content/examining-rise-non-performing-assets-india
[18] Pacha Malyadri, S. Sirisha, “Comparative Study of Non-Performing Assets in Indian Banking Industry”
[19] Gülden Kaya Uyanık, Neşe Güler, “A study on Multiple Linear Regression Analysis”, (2013)
[20] “An Analysis of Agriculture Sector in Indian Economy” IOSR Journal of Humanities And Social Science (IOSR-JHSS) Volume 19, Issue 1, Ver. X (Feb. 2014), PP 47-54
[21] “Impact of Agricultural Inputs on Agricultural GDP in Indian Economy” by Tamma Koti Reddy, Madhubanti Dutta, Department of Economics, IBS Hyderabad (A Constituent of ICFAI Foundation for Higher Education), Hyderabad, India.
[22] “Estimation, Analysis and Projection of India’s GDP, A Time Series Model” by U. R Daga, Rituparna Das and Bhishma Mahesh
[23] “High Ratio of Agriculture NPAS In Priority Sector Lending by Public and Private Banks in India –Reasons, Suggestions and Discussions” by Arpita Baijal - International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.
[24] “Non-Performing Loans and Terms of Credit of Public Sector Banks in India: An Empirical Assessment’’ by Rajiv Ranjan and Sarat Chandra Dhal
[25] Abhini S, Feature Selection with sklearn and Pandas, Towards Data Science, 11 February 2019. https://towardsdatascience.com/feature-selection-with-pandas-e3690ad8504b
[26] RBI Working Paper Series No. 03 “Re-emerging Stress in the Asset Quality of Indian Banks: Macro-Financial Linkages” Shashidhar M. Lokare
[27] India Agricultural Trade: Expanding Export Opportunities Amid Persistent Limitations, USDA Foreign Agricultural Service 21 December 2016. https://www.fas.usda.gov/data/india-agricultural-trade-expanding-export-opportunities-amid-persistent-limitations