# Yield Forecast: Exploring Regression Techniques in Agriculture

**Submitted for**

**BUSINESS FORECASTING METHODS AND APPLICATIONS**

Submitted by:

**(502304196) Chirag Bhatia**

**(502304215) Vibhuti Jain**

Submitted to-

**Dr. Nitin Arvind Shelke**



**L M Thapar School of Management**

**Jan-May 2024**

**INDEX**



| | Sr No | Content | Page No | | --- | --- | --- | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

| **LIST OF FIGURES**   | **Figure No.** | **Title** | **Page No.** | | --- | --- | --- | | 2.1 | Block diagram | 12 | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  |     **LIST OF TABLES**   | **Table No.** | **Title** | **Page No.** | | --- | --- | --- | | **1.1** |  | 8 | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  |   **Note: List of Figures and List of Tables shall be on separate pages.**  **ABSTRACT**  This report delves into the application of multiple regression techniques for business forecasting using agricultural data, with a focus on predicting rice yield. The study centers on Chhattisgarh (State Code 14) to streamline analysis. By employing multiple linear regression, k-nearest neighbors regression, support vector regression, and polynomial regression, we aimed to determine the most effective model. Results showcase the efficacy of polynomial regression, achieving an impressive accuracy score of 97.82, followed closely by multiple linear regression at 93.4. K-nearest neighbors regression scored 83.515, while support vector regression attained 92.47. These findings provide valuable insights for stakeholders in the agriculture sector, demonstrating the potential of regression techniques for accurate forecasting. |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **INTRODUCTION AND RELATED WORK**  In contemporary agriculture, the ability to predict crop yields accurately is crucial for effective resource management and decision-making. This study focuses on the utilization of regression techniques to forecast rice yield, a vital staple crop, within the context of Chhattisgarh (State Code 14), India.  The primary aim of this work is to evaluate the effectiveness of multiple regression methods, including multiple linear regression, k-nearest neighbors regression, support vector regression, and polynomial regression, in predicting rice yield. By applying these techniques to agricultural data encompassing factors such as district code, year, and area under rice cultivation, we seek to provide actionable insights for stakeholders in the agricultural sector.  The relevance of this problem lies in its implications for food security, economic stability, and sustainable agricultural practices. Accurate yield predictions enable farmers, policymakers, and agricultural organizations to plan cultivation strategies, allocate resources efficiently, and mitigate risks associated with fluctuating yields.  Through this study, we endeavor to contribute to the growing body of research on agricultural forecasting, aiming to develop robust models that can aid in decision-making processes and ultimately enhance the resilience and productivity of agricultural systems.  **SOFTWARE USED**  **For conducting the analysis and implementing the regression models, Jupyter Notebook was the primary software utilized. Jupyter Notebook provides an interactive computing environment that enables the creation and sharing of documents containing live code, equations, visualizations, and narrative text.**    **METHODOLOGY**  DATA PRE-PROCESSING Under the pre-processing phase, minimal cleaning was necessary due to the pristine nature of the dataset. No duplicate values or null entries were detected, signifying the dataset's cleanliness and integrity. This absence of data anomalies streamlined the pre-processing workflow, allowing for a more focused effort on feature selection and transformation tasks essential for regression model training.  Renamed Columns  Filter For State Code 14    Handling Categorical Data    Extracting the Dependent and the Independent Variable      Multiple Linear Regression  K-Nearest Regression    SVR Model    Polynomial Regression    **EXPERIMENTAL RESULTS**  Multiple Linear Regression -  We used a method called multiple linear regression to guess how much rice would be produced in 1966. We picked some simple numbers for the rice fields - 548 for the area and 185 for the production, which are already in the dataset. Our guess was that 388.632 units of rice would be produced. When we checked the actual data, we found that 337.42 units were produced. Even though the numbers weren't exactly the same, they were pretty close. We also used a score called R-squared, which told us that our guess was about 93.4% accurate. This means our method is pretty good at predicting rice yield based on the information we have.    K Nearest Regressor  We used K nearest regressor to estimate how much rice would be produced in 1970 in Durg district. We used the rice area (571.60) and rice production (473.60) for this prediction, which are already in the dataset. Our estimate was 760.88 units of rice, but when we checked the real data, we found that 828.55 units were actually produced. This means our estimate was a bit off. The R-squared score, which tells us how accurate our estimate is, came out to be 83.51%. While this score shows some level of accuracy, it's not as high as we'd like it to be.    SVR  With Support Vector Regression, we forecasted the rice yield for the year 2017 in Durg district, using a rice area of 554 and rice production of 430. Our prediction suggested a yield of 881.54 units. Although we couldn't compare this prediction with real-time data, our confidence in its accuracy is bolstered by an R-squared score of 92.47%. This high score indicates that our model's estimates are likely to be very close to the actual yield.    Polynomial Regression  Despite our attempts, we encountered an error when attempting to forecast using the polynomial regression method. However, during our analysis, we found that the R-squared score for this method was 97.4%, which was the highest among the four regression techniques employed. This high score indicates that the polynomial regression model had the best performance in explaining the variance in the data, despite the technical challenges encountered during the forecasting process.    **CONCLUSION**  In summary, our investigation into different ways of predicting rice yield in agriculture has provided some interesting findings. We tested four methods: multiple linear regression, k-nearest neighbors regression, support vector regression, and polynomial regression.  Our results showed that multiple linear regression gave us pretty accurate predictions, getting close to the actual yields with an accuracy score of 93.4%. K-nearest neighbors regression was okay but not as good, scoring 83.51%. Support vector regression did well too, with a score of 92.47%.  Surprisingly, even though we had trouble using polynomial regression, it turned out to be the most accurate method with a score of 97.4%. This tells us that it's essential to look at both how well a method works and how easy it is to use in real situations.  Overall, our study shows that these regression methods can be really helpful in predicting rice yields, which is great news for farmers and others involved in agriculture. By continuing to research and experiment with these methods, we can make even better predictions, helping to improve farming practices and food production.  **REFERENCES**  **GitHub Repository Link** |  |
| Minimum 5 reference is required (You may use the research article) |  |

**Typesetting:**

* Font: Times new roman
* Font size for the text: 12, Font size for Chapter name: 16 (Bold)
* All material should be typed in 1.5 line spacing using times new roman and the vertical space between paragraphs shall be 2.5 line spacing.
* The recommended margins are 25 mm (1 inch) for top, bottom, right and left with an extra 13 mm (0.5 inch) for binding on the left. Other than page numbers, no material should intrude into these margins.

Note: Plagiarism should not be more than 20%