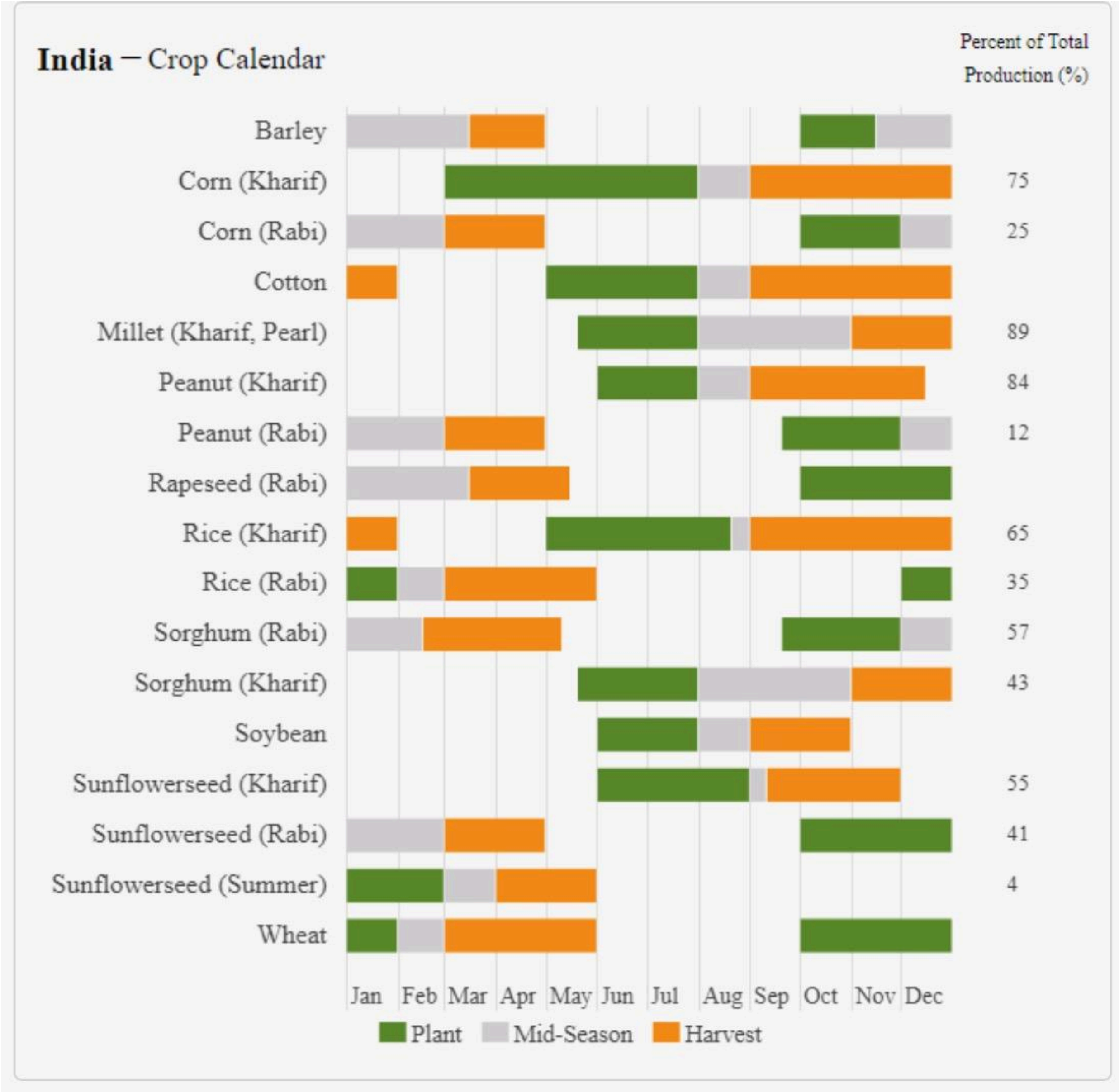
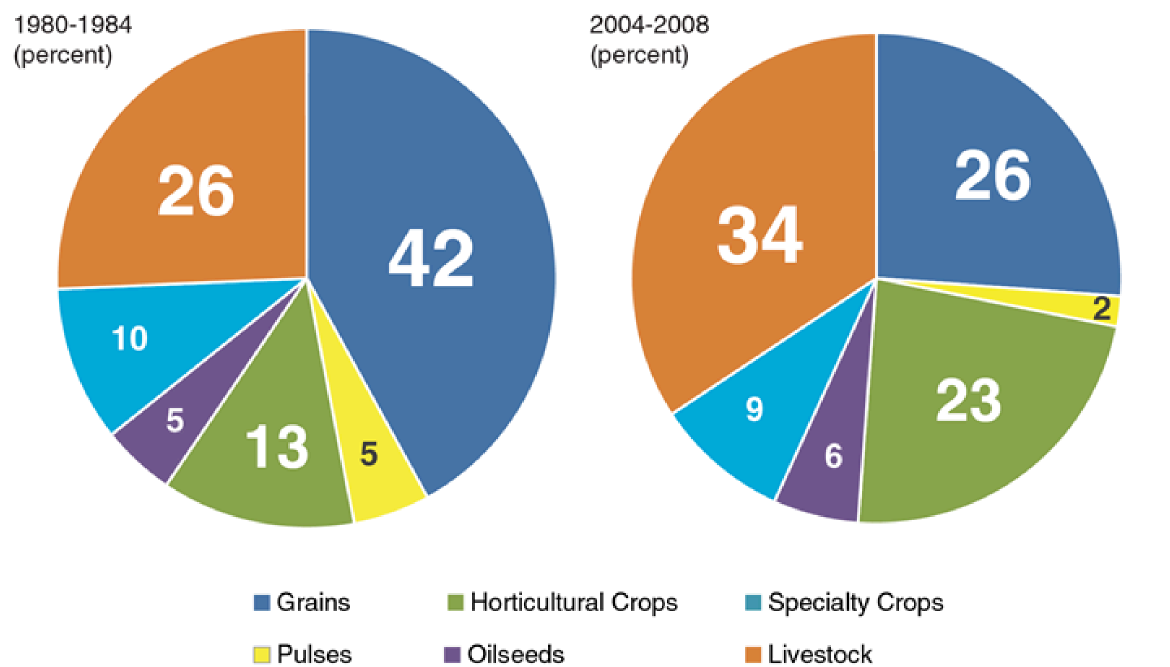
India stands as the world's largest country by population and seventh largest country by area, a factor that inherently underscores the critical need for sustainable agricultural practices to ensure food security for its billions of inhabitants. The sustainability of people's daily food source is not just a matter of national concern but has global implications, considering India's significant role in the global food market. As a major producer of various staple crops, including rice, wheat, pulses, any fluctuation in India's agricultural output can have far-reaching effects on global food prices and availability. This project is aimed at predicting crop production in India, hoping to leverage machine learning techniques to tackle the challenge of optimizing agricultural output. By analyzing various features such as season, geographical data, and crop variety, the project seeks to explore and forecast crop production. This predictive goal offers the potential to improve how agricultural policy is formulated and how resources are allocated. This project is interesting due to its unique and interdisciplinary structure, which merges the realms of agriculture, environmental science, and technology. It embodies the essence of innovation, applying the latest advancements in data analytics and machine learning to solve real-world problems that could potentially affect millions of lives. Moreover, by predicting crop production, the project provides insights that can help in mitigating the effects of potential food shortages, reducing waste, and enhancing food distribution strategies.By understanding the variables that impact crop production, the project can guide farmers towards more sustainable agricultural practices, reducing the environmental footprint of farming and ensuring the long-term viability of India's agricultural sector.

The dataset stands out for its completeness. This characteristic is vital as it ensures that the dataset provides a comprehensive overview of the variables influencing crop production in India. A well-constructed description accompanying the dataset offers clear insights into what the data represents. Credibility is another cornerstone of the dataset's selection. The source of the original data is transparently indicated, assuring of its authenticity and reliability. By relying on a dataset whose origins are clear and verifiable, the project establishes a solid foundation that readers can trust. Compatibility is equally critical. The dataset is structured in such a way that there are no ambiguous columns or undefined values, such that each piece of data can be accurately interpreted and utilized in predictive models. The cleanliness and usability of the dataset is also another key component. With minimal null values and a well-maintained structure, the dataset is primed for analysis without the need for extensive cleaning or preprocessing. In summary, the dataset for predicting crop production in India was chosen due to its completeness, credibility, compatibility, and cleanliness. These attributes ensure that the dataset is not only easy to work with but also robust and reliable, providing a strong foundation for generating accurate and meaningful predictions.



https://ipad.fas.usda.gov/countrysummary/Default.aspx?id=IN

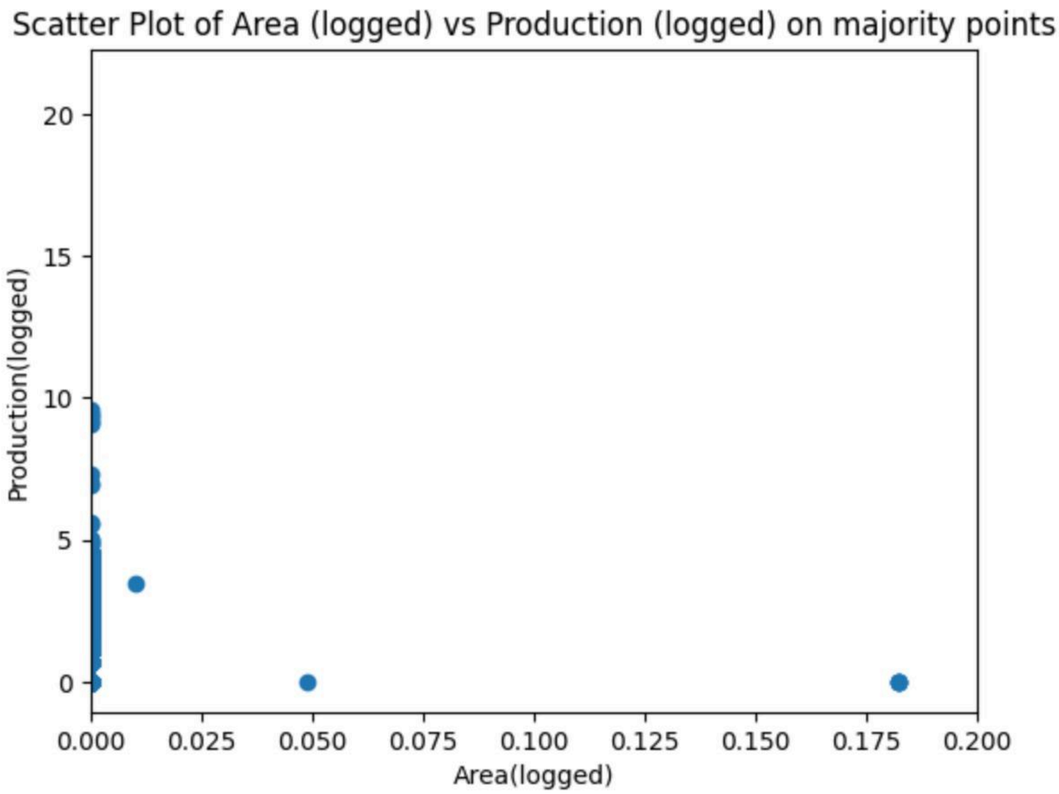
Agricultural output shares from livestock and horticultural crops in India

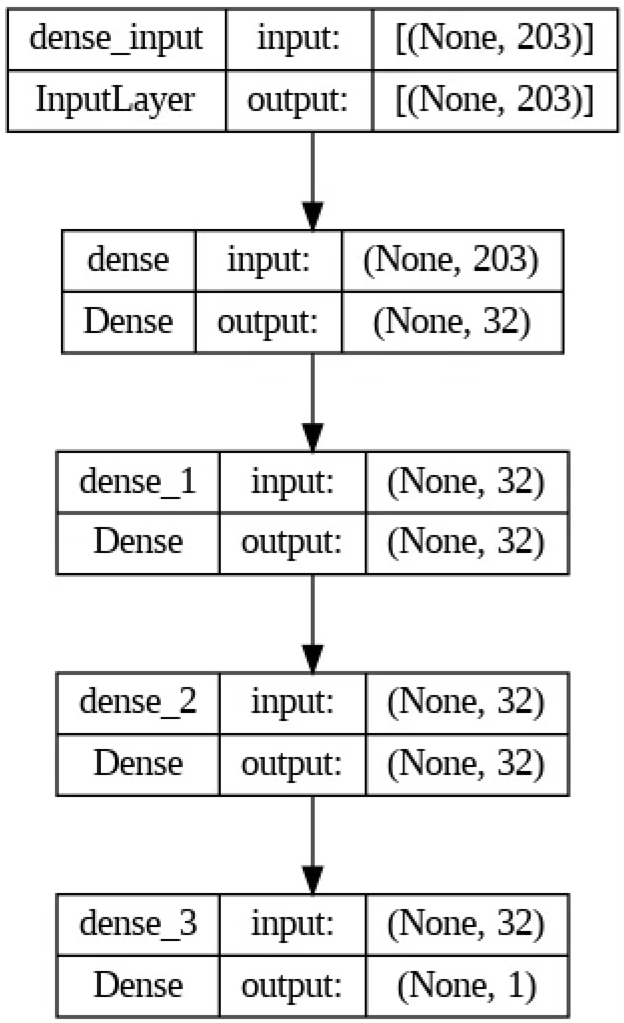


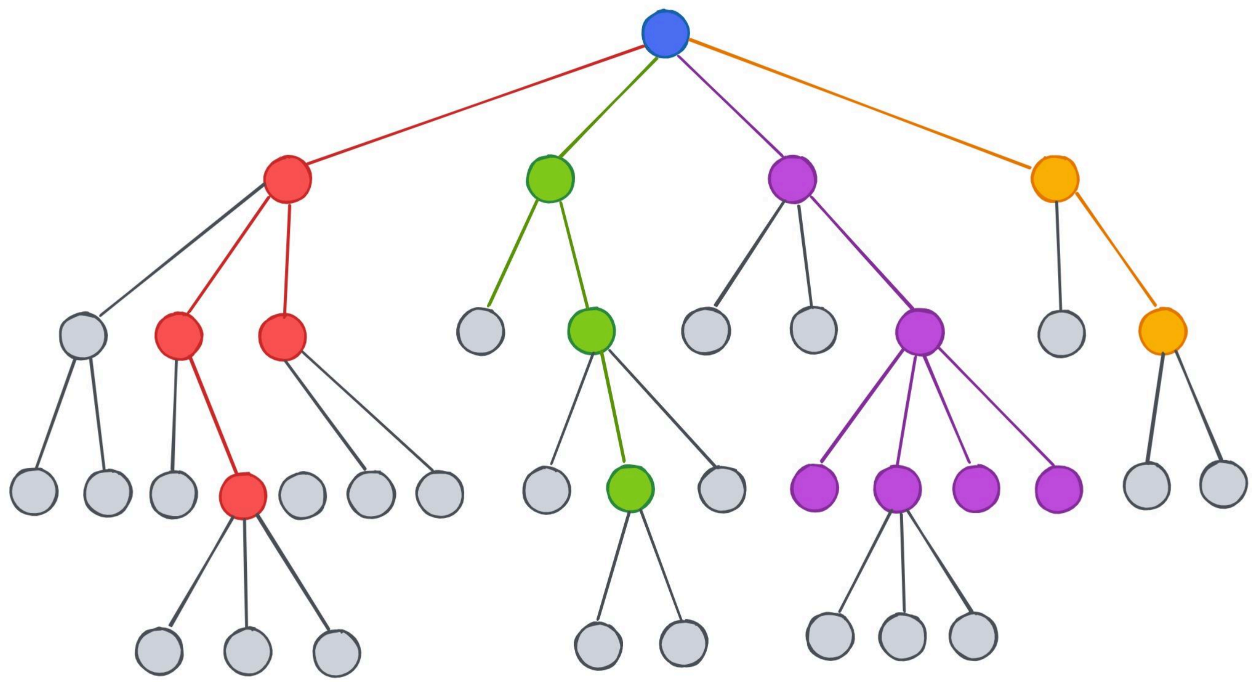
Note: Numbers in the figures reflect percentage shares of output. Shares may not sum to 100 because of rounding.
Source: USDA Economic Research Service estimates.

https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=78652

$$y = \beta_0 + \sum \beta_{\text{crop}} X_{\text{crop}} + \sum \beta_{\text{season}} X_{\text{season}} + \sum \beta_{\text{crop\_year}} X_{\text{crop\_year}} + \sum \beta_{\text{state}} X_{\text{state}} + \sum \beta_{\text{season\_crop}} X_{\text{season\_crop}} + \sum \beta_{\text{cropYear\_season}} X_{\text{cropYear\_season}} + \epsilon$$


Scatter Plot of Area (logged) vs Production (logged) on majority points
Production(logged)
Area(logged)
0
5
10
15
20
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200


dense_input
input:
[(None, 203)]
InputLayer
output:
[(None, 203)]
dense
input:
(None, 203)
Dense
output:
(None, 32)
dense_1
input:
(None, 32)
Dense
output:
(None, 32)
dense_2
input:
(None, 32)
Dense
output:
(None, 32)
dense_3
input:
(None, 32)
Dense
output:
(None, 1)



https://botpenguin.com/glossary/decision-trees