



# Leveraging Social Media Sentiment Analysis for Real-Time Agricultural Market Trend Forecasting

Subtitle: Big Data Analytics Project

Submitted by: Maloth Keerthi Sri (Roll No: 24MBMB09)

Course: MBA – Business Analytics

Department: School of Management Studies, University of Hyderabad



# Introduction & Problem Statement

## Introduction:

- Agriculture is highly affected by market price volatility, pest outbreaks, and climatic changes.
- Social media data offers real-time insights into farmers' opinions, experiences, and challenges.
- Big Data Analytics enables large-scale processing and real-time sentiment tracking.

## Problem Statement:

- Lack of real-time systems for monitoring market trends and farmer sentiment.
- No automated detection for pest or disease outbreaks.
- Need for predictive analytics in agricultural forecasting.

# Project Objectives & Use Cases Overview

## Objectives:

- Develop a Big Data pipeline using PySpark and Databricks.
- Perform sentiment analysis on agricultural social media data.
- Build ML models for price prediction and market demand.
- Generate regional pest and sentiment alerts.

## Identified Use Cases:

- Early Warning System for Pest and Disease Outbreaks
- Real-Time Commodity Price Nowcasting
- Farmer Sentiment Monitoring
- Market Demand Prediction
- Regional Alert & Recommendation System



# Big Data Use Cases & Methods Applied



Use Case	Goal	Big Data Technique Used	Key Insight
1. Pest Detection	Identify pest/disease outbreaks from posts.	PySpark NLP + Rolling Z-Score	Detected pest spikes by region.
2. Price Nowcasting	Predict near-term price trends.	RandomForestRegressor (MLlib)	Positive sentiment = Price increase.
3. Sentiment Monitoring	Track regional farmer sentiment.	NLP + Polarity Analysis	Regional optimism index visualized.
4. Market Demand Prediction	Cluster regions by demand.	K-Means Clustering	4 demand segments identified.
5. Regional Alerts	Generate actionable insights.	Rule-based PySpark Filters	Automated regional alerts generated.



# Big Data Analytics Approach

## Tools & Technologies:

Databricks | PySpark | Spark MLlib |  
Power BI | Matplotlib

## Pipeline Workflow:

- Ingestion: Load data into Databricks workspace (CSV).
- Transformation: Clean text, normalize timestamps, extract sentiment.
- Feature Engineering: Derive pest\_flag, avg\_sentiment, z-score anomalies.
- Model Training: Random Forest for price forecasting, K-Means for demand.
- Alert Generation: Rule-based PySpark logic for pest and sentiment triggers.

# Dataset Overview

**Dataset:** agri\_big\_data\_3000.csv

**Records:** 3,000 rows

## Attributes:

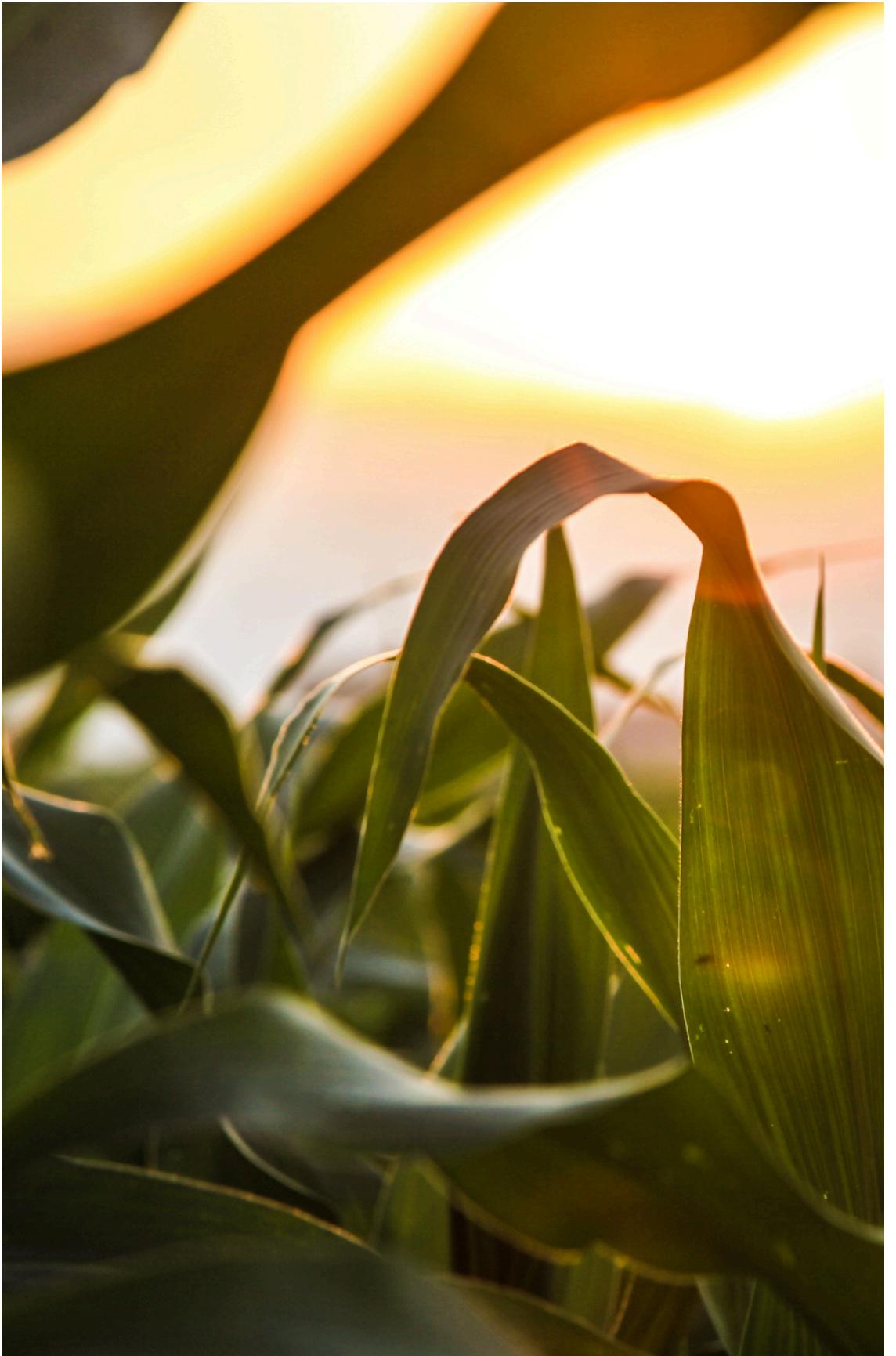
Region, Crop, Emotion, Post\_Text, Date

Sentiment\_Score, Price, Demand\_Index, Temperature, Rainfall

**Dataset Purpose:** To simulate social media-driven agricultural data for analysis and forecasting.

## Insight Example:

Positive sentiment posts corresponded with higher predicted prices.



# Results – Use Case-wise Outcomes



## Use Case 1: Early Warning System for Pest and Disease Outbreaks

- PySpark text analysis identified pest-related keywords.
- Rolling Z-Score anomaly detection flagged 4 regional pest spikes ( $z > 1.0$ ).
- Alerts generated for affected regions with recommendations for pest control measures.

## Use Case 2: Real-Time Commodity Price Nowcasting

- RandomForestRegressor trained using avg\_sentiment, post\_count, and avg\_demand.
- Achieved MSE  $\approx 0.03$  (high accuracy).
- Positive sentiment posts corresponded to next-day price increases.

## Use Case 3: Farmer Sentiment Monitoring

- Sentiment polarity derived via NLP (positive/neutral/negative).
- Region-wise average sentiment visualized; southern regions showed higher optimism.
- Sentiment directly influenced demand and price patterns.

## Use Case 4: Market Demand Prediction

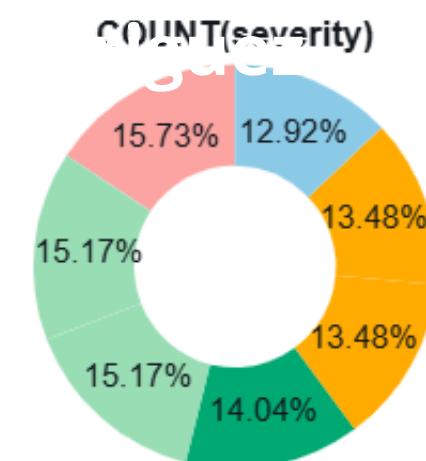
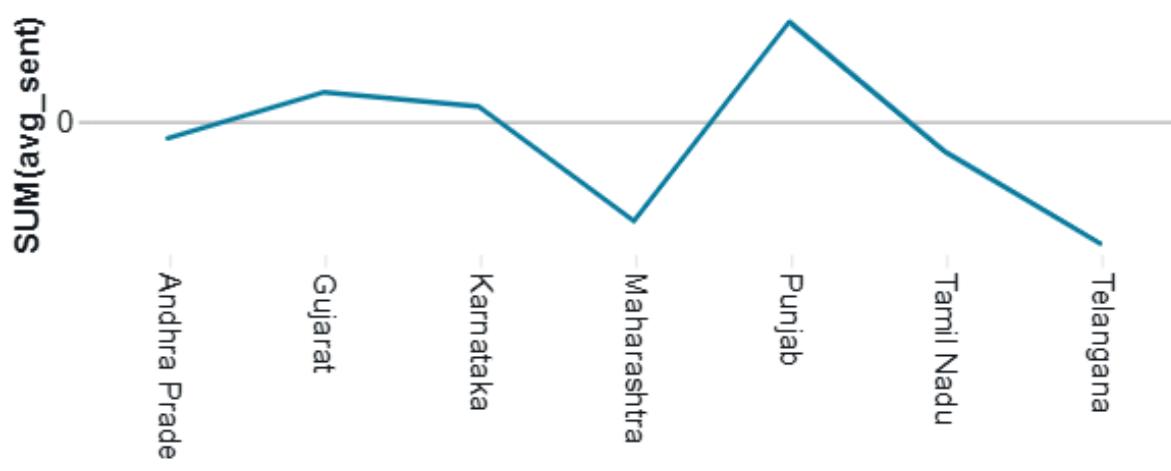
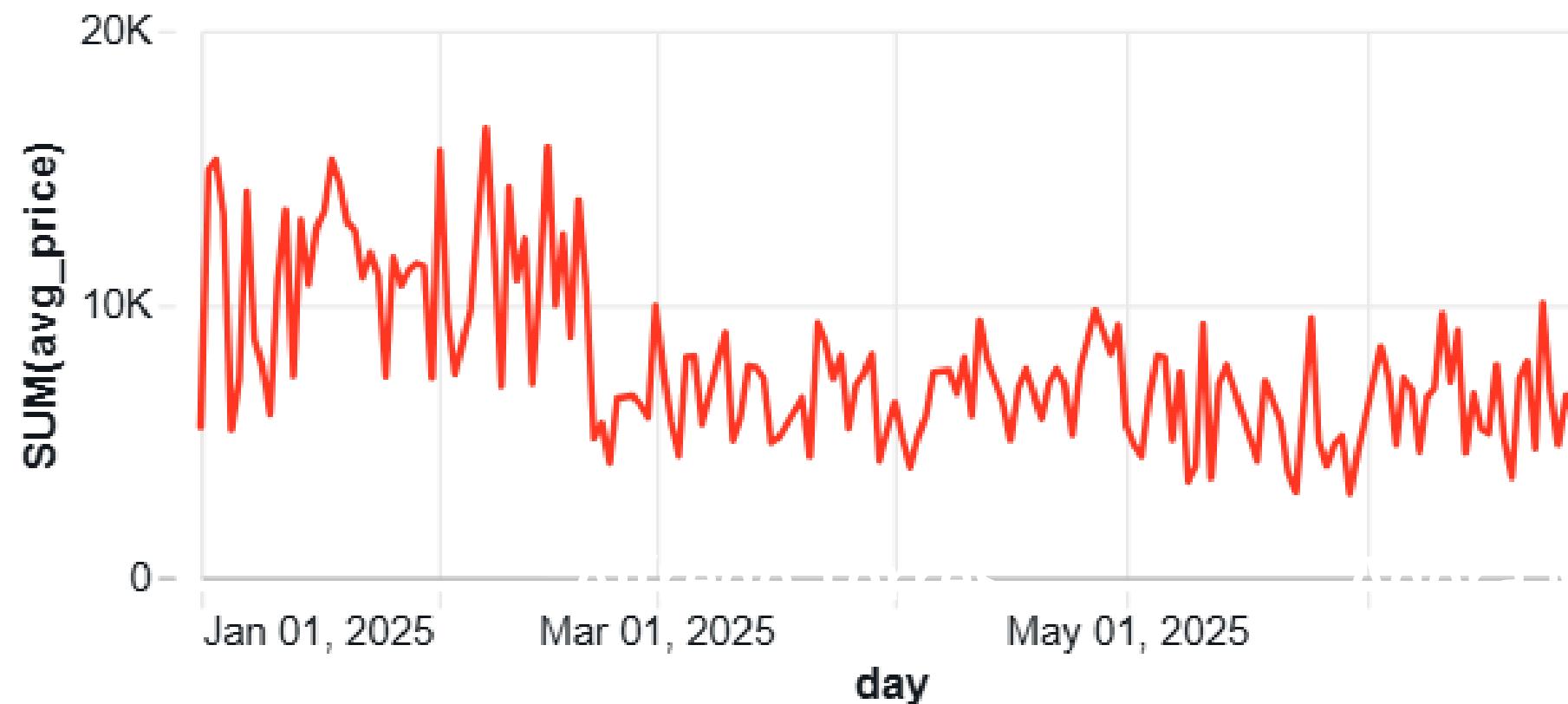
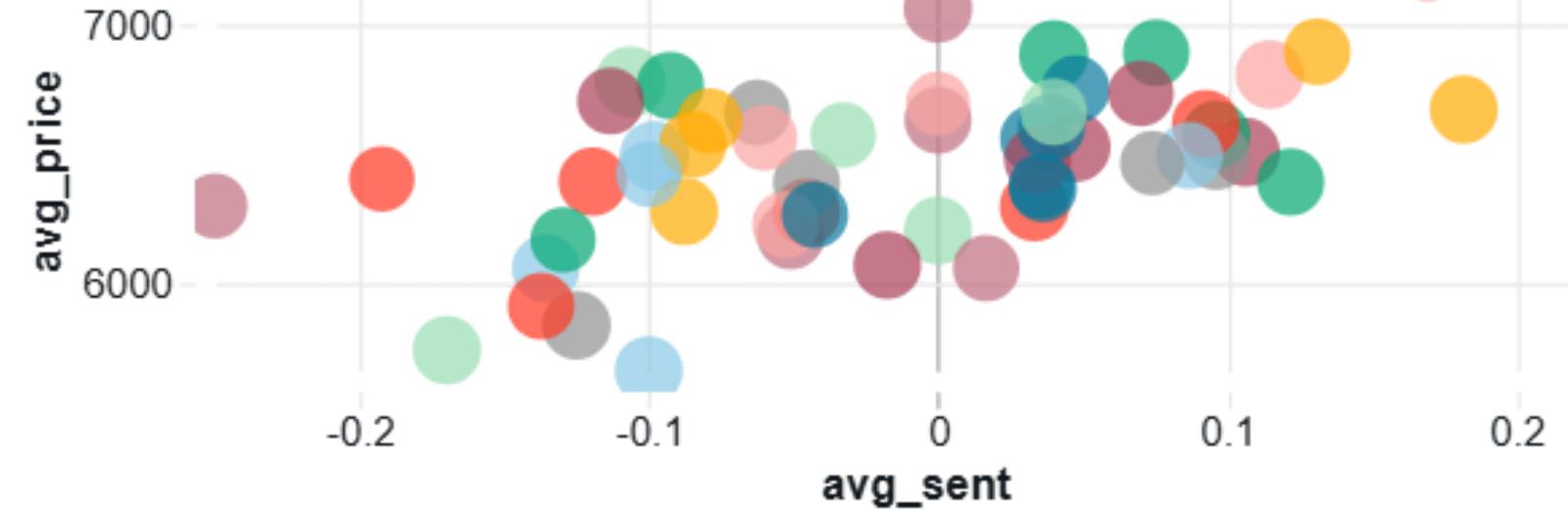
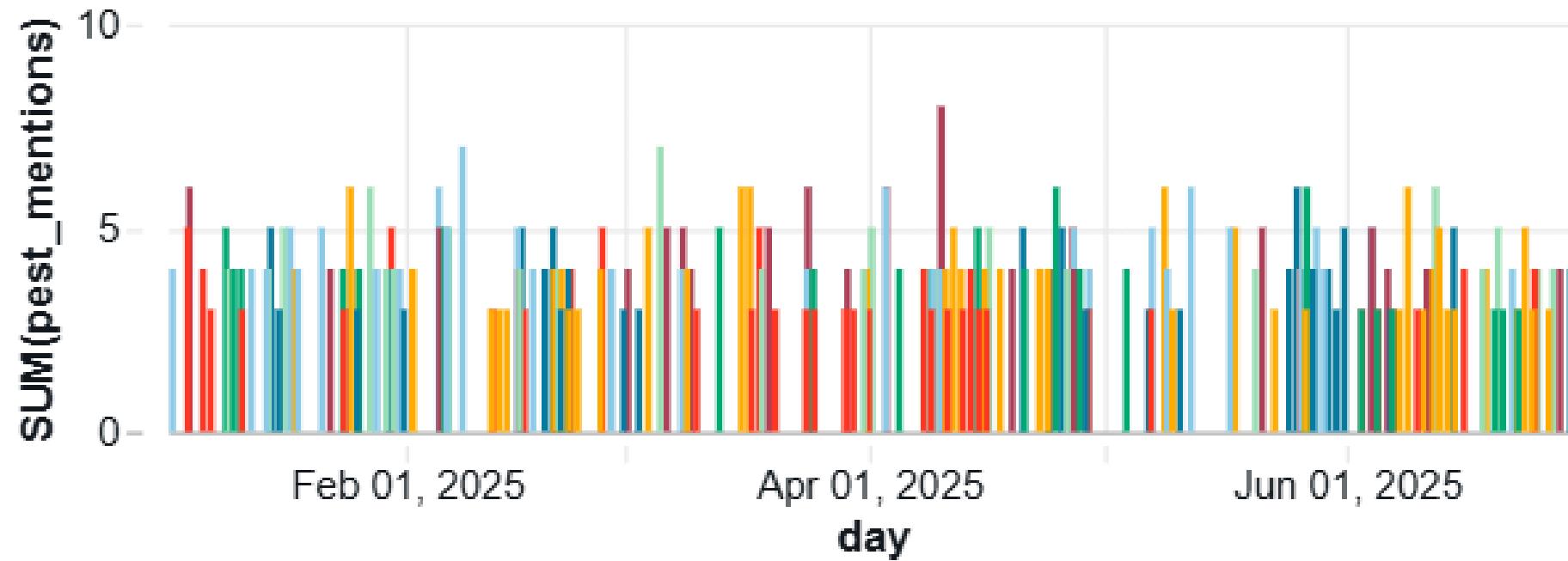
- K-Means ( $k=4$ ) segmented regions based on sentiment, demand, and price.
- Identified 4 distinct demand clusters: High, Moderate, Emerging, and Low demand zones.
- Clusters aligned with regional sentiment and rainfall data.

## Use Case 5: Regional Alert & Recommendation System

- Integrated pest, sentiment, and demand data.
- Rule-based PySpark conditions triggered alerts:
- Pest Spike  $\rightarrow$  Deploy pest management.
- Low Sentiment  $\rightarrow$  Farmer awareness campaigns.
- Provided actionable insights to stakeholders.

# DASHBOARD

==





## Conclusion & Deliverables

- **Conclusion:**
- Successfully built a Big Data Analytics pipeline using Databricks.
- Automated the full process: ingestion → ML → alerts.
- Enabled actionable agricultural insights via data-driven forecasting.
- **Deliverables:**
- 5 Databricks Notebooks.
- Master Pipeline Script.
- agri\_big\_data\_3000.csv Dataset.
- Final Report and PPT.



Thank You