BIG DATA APPLICATIONS IN MODERN FARMING

Subject: Big Data Analytics
Submitted by: Ankit Singh (Roll No: 24MBMA23)
Course: MBA (Gen.)
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.
- Big Data Analytics enables large-scale processing and real-time sentiment tracking.
- Digital Data from farmers and online platforms helps track their challenges

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:

- Pest and Disease Warning for farmers
- Real-Time Crop Price Predictions
- Checking Farmers' Sentiment
- Predicting Market Demand
- Local Farming Alerts and Updates

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: Pest and Disease Warning for Farmers

- 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 Crop Price Predictions

- RandomForestRegressor trained using avg_sentiment, post_count, and avg_demand.
- Achieved MSE ≈ 0.03 (high accuracy).
- Positive sentiment posts were associated with next-day price increases.

Use Case 3: Checking Farmers' Sentiment

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

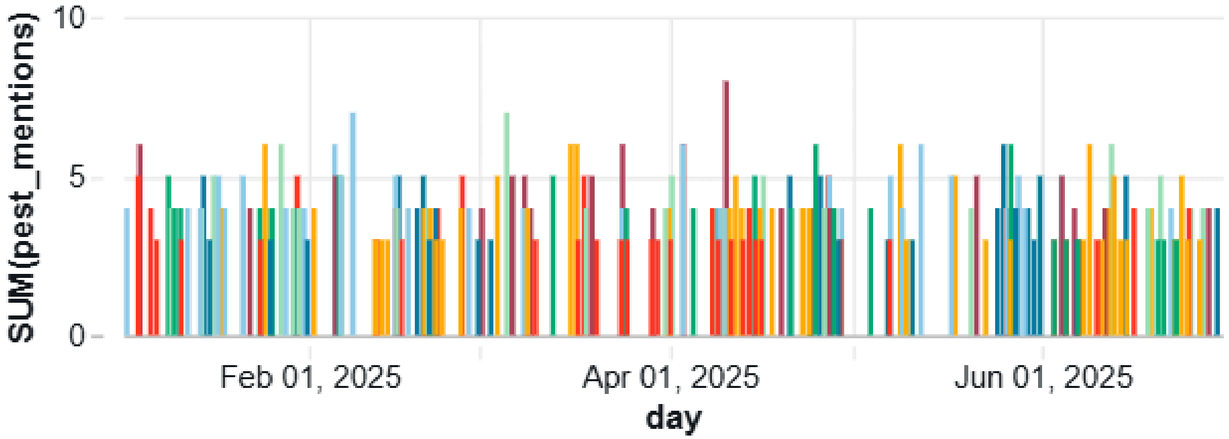
Use Case 4: Predicting Market Demand

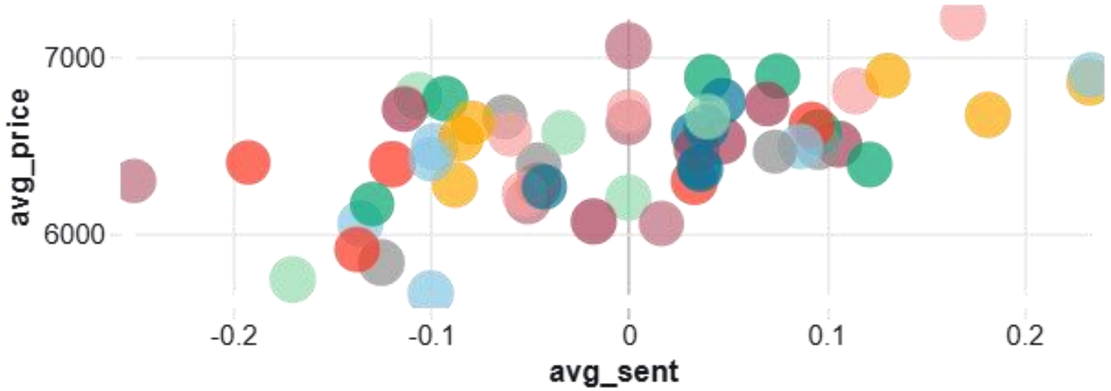
- K-Means ($k=4$) segmented regions based on sentiment, demand, and price.
- Identified 4 distinct demand clusters: High, Moderate, Emerging, and Low demand zones.
- Clusters aligned with regional sentiment and rainfall data.

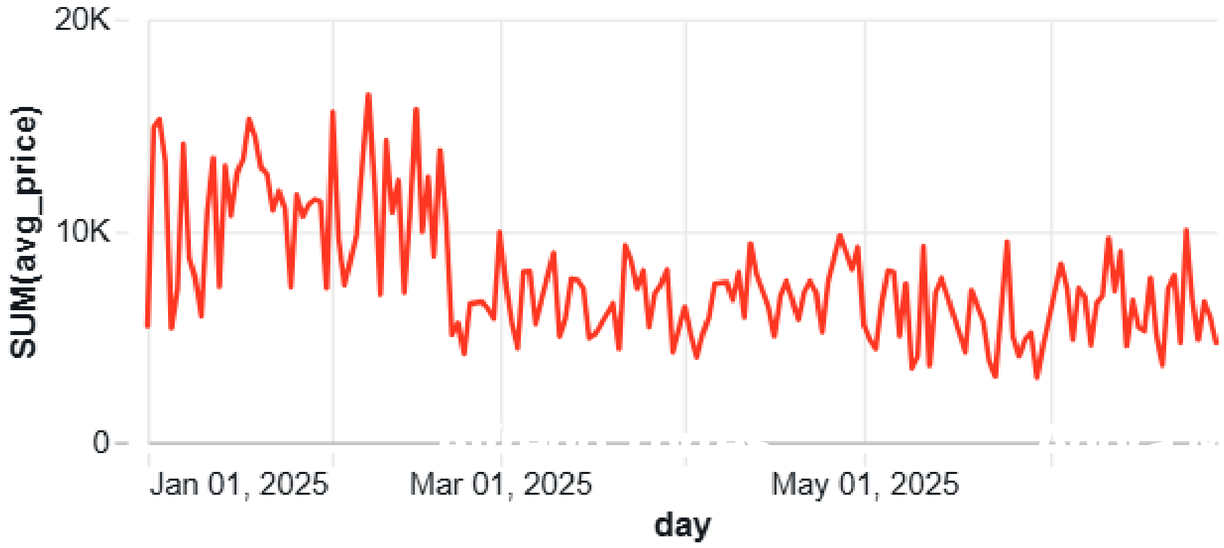
Use Case 5: Local Farming Alerts and Updates

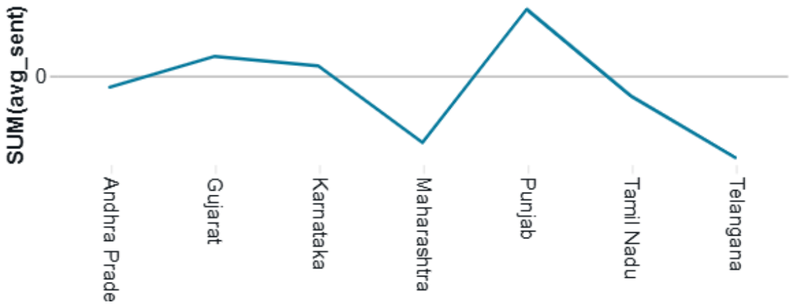
- Integrated pest, sentiment, and demand data.
- Rule-based PySpark conditions triggered alerts:
- Pest Spike $\rightarrow$Deploy pest management.
- Low Sentiment $\rightarrow$Farmer awareness campaigns.
- Provided actionable insights to stakeholders.

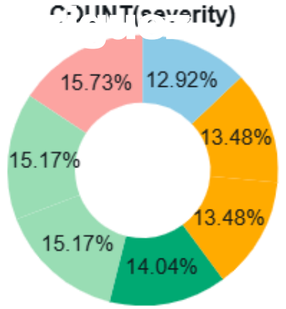
DASHBOARD


SUM(pest_mentions)
10
5
0
Feb 01, 2025
Apr 01, 2025
Jun 01, 2025
day


avg_price
7000
6000
-0.2
-0.1
0
0.1
0.2
avg_sent


SUM(avg_price)
20K
10K
0
Jan 01, 2025
Mar 01, 2025
May 01, 2025
day


SUM(avg_sent)
0
Andhra Prade
Gujarat
Karnataka
Maharashtra
Punjab
Tamil Nadu
Telangana


15.73%
12.92%
13.48%
15.17%
13.48%
15.17%
14.04%

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