

FraudDetectX – Detecting Fake Subsidy Claims Using Big Data

FraudDetectX is a computer program project that assists the government in identifying bogus or suspicious subsidy claims through big data technologies and machine learning. We use only software – no IoT or hardware. Our project scans big volumes of subsidy data and identifies records that appear to be fraud (such as claiming an excess amount of money or producing false Aadhaar).

Why This Project?

Each year, the government distributes subsidies on farming, LPG, education, and others. But some individuals exploit this by:

- Filing duplicate claims
- Providing incorrect data
- Claiming excess money

This results in massive loss of money. Our project catches these forgeries early through technology.

What Our Project Does

1. We accept actual or dummy subsidy data (JSON or CSV format).
2. The data is then stored in Hadoop HDFS – a big data storage repository.
3. We process the data using MapReduce to remove missing or incorrect entries.
4. The clean data is passed to Apache Spark to be processed quickly.
5. With Spark MLlib, we execute a machine learning algorithm to identify anomalies (suspicious cases).
6. The system displays a list of claims that appear to be fraud (e.g., suspiciously high value or multiple claims from the same Aadhaar).

Tools & Technologies Used

- Hadoop HDFS – maintains large data files
- MapReduce – runs the data in batches
- Apache Spark – does fast calculations
- MLlib (Spark's ML library) – applies machine learning
- Python or Scala – programming languages used

How We'll Show This Project (Demo)

We will:

1. Upload a sample dataset of subsidy claims
2. Show how the system processes the data using Hadoop and Spark
3. Show how frauds are detected (like 1 person submitting 5 times)
4. Explain results in a simple table – showing which entries were flagged

Expected Outcome

The system will highlight:

- Duplicate Aadhaar numbers
- Very high amounts of claims
- Individuals making claims excessively

It makes it easy for authorities to detect suspicious activity.

Conclusion

FraudDetectX is an intelligent, scalable, and software-based system to assist in the detection of subsidy frauds. It employs Hadoop and Spark to deal with big data, and ML to take informed decisions. It has the ability to save crores of rupees and make subsidy systems transparent and equitable.