911 CALL SERVICE ANALYSIS AND DETECTION

By

Team PLC

AGENDA

- INTRODUCTION
- PROBLEM STATEMENT
- OBJECTIVE
- DATA COLLECTION
- PROJECT IMPLEMENTATION
- WHY PYSPARK?
- WORKFLOW
- BENEFITS
- CHALLENGES
- CONCLUSION

INTRODUCTION

- Emergency services are a foundation of social safety, offering quick assistance in times of need. As the demand for these services grows, so does the difficulty of controlling and improving their delivery.
- The capacity to respond to emergency calls promptly and efficiently can be the difference between life and death, and improving these services is an essential goal.
- Here, in this project we have taken priority as the main column to detect what type of priority the incoming call can have, which can be very helpful when there are emergencies in the received call.

PROBLEM STATEMENT

The problem statement of our project is to analyse the 911 call services data and detect the priority of the incoming call.

OBJECTIVE

- Analysing 911 call service data
- Finding which part of the location receives high-priority calls, and so on.
- Detecting the priority of the incoming calls.
- Usage of technologies such as Spark RDD, PySpark, and Spark SQL.

WHY USE PYSPARK?

- PySpark handles Big Data and helps to perform complex computations quickly.
- Has a rich set of APIs for processing, manipulating, and analysing data, which can be very helpful for data science tasks.
- It also has an MLlib library which can be very helpful in supporting machine learning algorithms.
- It can also be easily integrated with other tools of big data.

DATASET

- We have taken the dataset from https://data.gov/ website. This website has various open datasets of the U.S. government.
- The dataset which we have collected has 3161012 rows and 21 columns with a total size of 692 MB.

PROJECT IMPLEMENTATION

Technologies which we have used in this project are :

- PySpark
- Spark RDD
- Spark SQL

THE WORKFLOW OF OUR PROJECT

We have done our project in the following steps:-

- Analyse the data using both Spark RDD and Spark SQL
  - Get the description and summary of the dataset
  - Get the value counts of the columns using Spark RDD (map and reduceByKey) and Spark SQL.
  - Analyse the combination of target (priority) with other columns (district, description, Neighborhood, police district, police post, council district, sheriff districts, ZipCode).
  - Use Spark SQL to analyze the data using functions orderBy, asc, desc, and so on.

THE WORKFLOW OF OUR PROJECT- (Contd.)

- Preprocessing the dataset which included the following :-
  - Data Cleaning - Handle missing values by removing the null data.
  - Feature Selection
    - Drop unnecessary columns (recordId, callKey, callNumber, incidentLocation, location, Community_Statistical_Areas, VRIZones, NeedSync, IsDeleted, ESRI_OID).

THE WORKFLOW OF OUR PROJECT(Contd.)

- Data Transformation
  - Handle the categorical columns and also the target column using StringIndexer and Vector Assembler.
- Train and Test Split - Split the dataset into train and test with a proportion of 80 and 20 respectively.
- Modeling
  - Implement the decision tree classifier with maxBins of 4000
  - Get the visualization of the decision tree.
  - Get the feature importances of the model.
  - Transform the model for the test dataset.
- Evaluate the model.

BENEFITS OF THE PROJECT

- Can help to detect the priority of incoming calls which can help in concentrating on emergency and high-priority calls instead of concentrating on unnecessary calls.
- Fastest implementation of the project because of using PySpark.
- Can help in taking preventive steps based on identified patterns.

CHALLENGES

- Complexity in implementing the code.
- Issues in the dataset values.

CONCLUSION

- In recent days we have been watching a high volume of crimes being registered.
- It is important to have a call service to overcome these crimes which is done by 911 call services.
- As the number of crimes increases, the incoming calls for the 911 service also increase.
- Analysing the data and detecting the priority is useful. Detecting the priority of the incoming call will help to give a fast and efficient solution to the crimes which has been done in our project.

ANY QUESTIONS?


Thank You!