

# Project Proposal

## Group 10

### Group:

Project Group 10

### Team Members:

- Monank Savaliya (012415260)
- Umang Kakaiya (012449515)
- Vishweshkumar Patel (012461371)

## 1. Description of the proposed project.

Title: Predicting “Call Type Group” for Fire Department Calls in city of San Francisco based on historical records from Year 2000 to Year 2018.

- The project focuses to use historical records from Year 2000 to Year 2018 of Fire Department Calls of city of San Francisco to predict(classify) “Call Type Group” for a specific location for certain day and time of week.
- The outcome of the project is to predict(classify) “Call Type Group” from four different types - Fire, Alarm, Potential Life Threatening and Non-Life Threatening for a call made to Fire Department of San Francisco city.
- The use-case will provide more insights to respective fire departments to take important decisions to improve their services and save more lives.

## 2. Dataset details:

Dataset Source Type:

- Public Dataset of “Fire Department Calls for Service” from DataSF

Dataset Source:

- DataSF - <https://data.sfgov.org/Public-Safety/Fire-Department-Calls-for-Service/nuek-vuh3/data>
- Data dictionary - [https://data.sfgov.org/api/views/nuek-vuh3/files/ddb7f3a9-0160-4f07-bb1e-2af744909294?download=true&filename=FIR-0002\\_DataDictionary\\_fire-calls-for-service.xlsx](https://data.sfgov.org/api/views/nuek-vuh3/files/ddb7f3a9-0160-4f07-bb1e-2af744909294?download=true&filename=FIR-0002_DataDictionary_fire-calls-for-service.xlsx)

Dataset Size:

- Rows: 4.61 Million
- Columns: 34
- Data: From Year 2000 to Year 2018
- Size in GBs: 1.56 GB

# Project Proposal

## Group 10

### 3. Technologies, Languages, Framework and Cloud Platforms:

- Language: Python
- Framework: Pandas, NumPy, Sklearn, Apache Hadoop, Apache Spark, Seaborn, Keras, Tensorflow
- Cloud Platform (For Jupyter notebook usage): Google Collaboratory

### 4. Proposed Techniques and Methods:

- Following steps can be followed to predict 'Category' of Call Type Group-
  - Data Pre-processing:
    - ◆ In Data Pre-processing step, null values – missing values and outliers that can adversely affect the predicted(classified) value will be taken care of. Data Sampling techniques may also be applied for large datasets.
  - Data Exploration:
    - ◆ In Data Exploration step, various trends for various categories of "Call Type Group" can be visualized (through various statistical measures) to understand importance of available features and hidden relations among them.
    - ◆ May be these relations and trends can be discussed with domain experts to understand problem at hand at deep level.
  - Feature Extraction and Engineering:
    - ◆ In Feature Extraction and Engineering step, from available features one can extract new features using dimensionality reduction techniques – PCA, SVD, etc OR feature engineering techniques - Date and time features, numeric to categorical mappings, grouping sparse classes, etc. that are more suitable for prediction task to achieve maximum accuracy.
  - Model(s) Training and Testing:
    - ◆ In Model(s) Training and Testing step, selected appropriate models for classification tasks – SVM, Decision Trees, KNN, Neural Network, etc. can be trained using transformed features obtained through feature engineering step. Dataset will be divided in Training and Testing datasets, then again Training data is divided in Training and Validation data.

# Project Proposal

## Group 10

- Model(s) Evaluation and Selection:
  - ◆ Based on performance of model(s) on Training, validation and testing dataset (unseen data), it will be evaluated and model that provides highest accuracy measure (f-score, roc curve, precision, recall and accuracy) in cross validation and testing dataset is selected for classification purpose.
- Knowledge mining and Visualizations:
  - In Knowledge mining and Visualizations step, based on predicted(classified) category of “Call Type Group” for a certain location on a certain day and time – Fire department can take certain actions and execute certain decisions to enhance their services and save more lives.

### 5. Contribution of each member:

- Each member will train one supervised learning model and will perform end to end tasks of Data Science life cycle phases from Data Pre-processing to Model building and evaluation.
- Based on Model’s accuracy, we will give justification for choosing a certain model from those three models and will perform knowledge mining and visualization phase on chosen model.