**Explanation Document**

**AIM:**

This project aimed to analyse the Medical Claim Fraud Detection to predict the false insurance claim

**Dataset Explanation :**

A) Inpatient Data

This data provides insights about the claims filed for those patients who are admitted in the hospitals. It also provides additional details like their admission and discharge dates and admit d diagnosis code.

B) Outpatient Data

This data provides details about the claims filed for those patients who visit hospitals and not admitted in it.

C) Beneficiary Details Data

This data contains beneficiary KYC details like health conditions, region they belong to etc

**Data Preprocessing and Feature Engineering:**

1. Merging the Dataset according to the common key features between all the dataset.
2. At first join the inpatient and outpatient dataset by left (outpatient) and right (inpatient)join.
3. Then Join the beneficiary data set on ‘BeneID’.
4. Finally merge the provider dataset based on ‘Provider’ feature.
5. Finding insight between the columns and applied some data concepts helps to create a new feature for effective prediction.
6. Applied SMOTE to balance the target variable.
7. Splitting the dataset train as 90% of data and test as 10% of data.

**Deep Learning Model:**

To achieve the effective prediction Deep neural network (DNN) model uesd with Hybrid Grey Wolf - Whale Optimization technique applied for best parameters.

**Results:**

The DNN model with best optimization parameters achieved the above 90% of accuracy. DNN model got the best results when compare with other Machine Learning model.