World tender data set processing using ML algorithms

Frist I just collect the raw data set from Kaggle on world tenders store it in the csv format to process the data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So for this, we use data pre-processing task in the machine learning algorithm the fetching raw data and drop the null for getting better accuracy after then I just use chat gpt4 to quality analysis the data set

Importing required modules like NumPy .It Python library is used for including any type of mathematical operation in the code. It is the fundamental package for scientific calculation in Python. It also supports to add large, multidimensional arrays and matrices. So, in Python, And also importing pandas The last library is the Pandas library, which is one of the most famous Python libraries and used for importing and managing the datasets. It is an open-source data manipulation and analysis library

SPLITTING THE DATASET INTO THE TRAINING SET AND TEST SET

In machine learning data pre-processing, we divide our dataset into a training set and test set. This is one of the crucial steps of data pre-processing as by doing this, we can enhance the performance of our machine learning model.

Suppose, if we have given training to our machine learning model by a dataset and we test it by a completely different dataset. Then, it will create difficulties for our model to understand the correlations between the models.

If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset. Splitting the data set in the ratio 20:80

Next process Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data

visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more.by importing library called as Sklearn to visulation process

First I just tried Knn algorithms but it don't give better accuracy because of huge data set soo I goo with random forest Therefore, to execute random forest algorithm null values have to be managed from the original raw data set. And another aspect is that data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in given dataset. Even though after this got soo many error because of the data set

ERROR AND SOLUTIONS :

Major error

1. the code to resolve the "Value Error: Input contains NaN, infinity or a value too large for dtype('float32')" error can vary depending on the specific circumstances.
2. func 'isfinite' not supported for the input types,(using numpy to resolve taht problem.To fix this error, you can convert the array to a float data type before applying the isfinite function
   due to no proper data set in the form of string cant able to conver it to float sooo dropping some columns to process the algorithm even though getting some errors