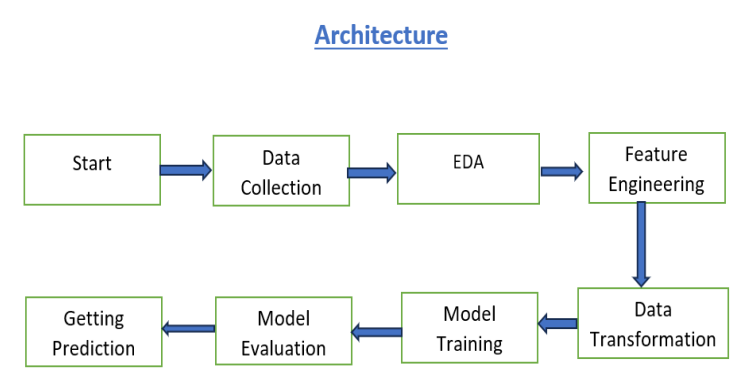
**Architecture Design**

*News Article Sorting*



**1. Data Description:**

Initially raw data is stored as csv file for the project which was provided by iNeuron

Columns:

▪ Text : News Article Text.

▪ Category : variable indicating whether the news is of business category, politics, tech, sports,entartainment etc.

**2. EDA**

• **Understand Data Structure:**

* Load the dataset and inspect the structure using functions like head(), info(), and describe().
* Check for data type, missing values, text preprocessing.

• **Visualize Data:**

* Use box plots, and pie plot to understand the distribution of features.
* Use bar plots and pie charts for categorical features.(news article text )

**3. Feature Engineering**

* **Handle Missing Values:**
  + Impute missing values using mean, median, mode, or other imputation techniques.
  + Consider removing features or rows with a high percentage of missing values.
* **Encode Categorical Variables:**
  + Apply one-hot encoding for nominal categorical features.
  + Use label encoding or ordinal encoding for ordinal features.
* **Scale and Normalize:**
  + Apply scaling techniques such as standardization (Z-score normalization) or min-max scaling.
  + Normalize features if necessary, especially for algorithms sensitive to feature scaling.

**4. Model Training:**

• **Split Data:**

* Divide the dataset into training and testing sets (e.g., 80/20 split) using functions like train\_test\_split.

• **Select and Train Models:**

* Choose appropriate machine learning algorithms based on the problem type (e.g., regression, classification). In this project our model is classification.
* Train models using training data (e.g., multinomial linear regression, decision tree classifier, SVC, neural networks).

**6. Model Validation**

• **Model Performance Metrics:**

* For classification: accuracy, precision, recall, F1-score, ROC-AUC.

• **Confusion Matrix:**

* For classification problems, use a confusion matrix to visualize true positives, false positives, true negatives, and false negatives.

**7. Getting Prediction**

• **Load Model:**

* Load the trained model and necessary preprocessing steps.

• **Preprocess Input Data:**

* Apply the same preprocessing steps to new input data as used during training.
* Ensure consistency in feature formats and encodings.

• **Generate Predictions:**

* Use the trained model to make predictions on new, unseen data.
* Output predictions in the desired format (e.g., class labels, continuous values).