**PROJECT TITLE : FAKE NEWS DETECTION USING NLP**

**PHASE 1:PROBLEM DEFENITION & DESIGN THINKING**

**PROBLEM DEFENITION:**

**THE PROBLEM IS TO DEVELOP A FAKE NEWS DETECTION MODEL USING A KAGGLE DATASET.WE CONSUME NEWS THROUGH SEVERAL MEDIUMS THROUGHOUT THE DAY IN OUR DAILY ROUTINE , BUT SOMETIMES IT BECOME DIFFICULT TO DECIDE WHICH ONE IS FAKE AND WHICH ONE IS AUTHENTIC.**

**SINCE ALL THE NEWS WE ENCOUNTER IN OUR DAY-TO-DAY LIFE IS NOT AUTHENTIC , HOW DO WE CATEGORIZE IF THE NEWS IS FAKE OR REAL? THE GOAL OF THIS PROJECT IS TO DISTINGUISH BETWEEN GENUINE AND FAKE NEWS BASED ON THEIR TITLES AND TEXT.**

**THIS PROJECT INVOLVES USING NATURAL LANGUAGE PROCESSING(NLP) TECHNIQUES TO PREPROCESS THE TEXT DATA , BUILDING A MACHINE LEARNING MODEL FOR CLASSIFICATION , AND EVALUATING THE MODEL’S PERFOMANCE.**

**DESIGN THINKING:**

1. **DATA COLLECTION : GATHER A DATASET OF NEWS ARTICLES LABELED AS EITHER REAL OR FAKE.**
2. **DATA PREPROCESSING : CLEAN AND PREPROCESS THE TEXT DATA. THIS INCLUDES REMOVING STOPWORDS, PUNCTUATION, AND SPECIAL CHARACTERS, AS WELL AS STEMMING OR LEMMATIZING WORDS.**
3. **FEATURE EXTRACTION : CONVERT THE TEXT DATA INTO NUMERICAL FEATURES. COMMON TECHNIQUES INCLUDE TF-IDF (TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY) AND WORD EMBEDDINGS LIKE WORD2VEC OR GLOVE.**
4. **MODEL SELECTION : CHOOSE AN APPROPRIATE MACHINE LEARNING OR DEEP LEARNING MODEL. COMMON CHOICES INCLUDE LOGISTIC REGRESSION, NAIVE BAYES, RANDOM FOREST, AND NEURAL NETWORKS LIKE LSTM OR BERT.**
5. **TRAINING : TRAIN THE SELECTED MODEL ON YOUR LABELED DATASET. ENSURE YOU SPLIT THE DATA INTO TRAINING AND TESTING SETS TO EVALUATE THE MODEL'S PERFORMANCE.**
6. **EVALUATION : USE METRICS LIKE ACCURACY, PRECISION, RECALL, AND F1-SCORE TO ASSESS HOW WELL THE MODEL CLASSIFIES FAKE AND REAL NEWS.**
7. **FINE-TUNING : ADJUST MODEL PARAMETERS AND FEATURES TO IMPROVE PERFORMANCE.**
8. **DEPLOYMENT : ONCE SATISFIED WITH THE MODEL'S PERFORMANCE, DEPLOY IT TO DETECT FAKE NEWS IN REAL-TIME OR BATCH PROCESSING.**
9. **MONITORING : CONTINUOUSLY MONITOR THE MODEL'S PERFORMANCE AND RETRAIN IT WITH NEW DATA TO ADAPT TO EVOLVING FAKE NEWS PATTERNS.**
10. **EXPLAINABILITY : CONSIDER TECHNIQUES TO EXPLAIN WHY THE MODEL CLASSIFIES A NEWS ARTICLE AS FAKE OR REAL. THIS HELPS BUILD TRUST IN THE MODEL'S DECISIONS**