

PROJECT DOCUMENT



**Student Name:** Charishma Maradapu

**Batch:** AI Elite 16



**Project Name:** Fake News Detection

**Domain:** Media

**Type of ML:** Supervised ML

**Type of Problem:** Classification

**Project Methodology:** CRISP-ML (Q)



**Phase I: Business and Data Understanding**

1. **Business Understanding:**Business Objective: Maximize usage of Application

Constraints: Minimize the Time complexity.  
  
Success Criteria:

1. *ML success criteria*: with average 96% accuracy score
2. Business Success criteria: -
3. Economy Success criteria: -
4. **Data Understanding:**

| **S No** | **Feature Name** | **Data Type** |
| --- | --- | --- |
| 1 | Title | Categorical |
| 2 | text | Categorical |
| 3 | Subject | Categorical |
| 4 | Date | Datetime |
| 5 | Label | Categorical |

**Phase 2: Data Preparation  
  
a) Exploratory Data Analysis:**

| **S No** | **Type** | **Feature Names** | **Observation** |
| --- | --- | --- | --- |
| 1 | Missing Values | - | - |
| 2 | Duplicates | Entire Data set | 209 |

**b) Data Cleaning/wrangling:**

| **S no** | **Type of Cleaning** | **Technique** | **Feature Name** | **Reason** |
| --- | --- | --- | --- | --- |
| 1 | Pre-processing | NLTK | text | Stop words, punctuations, word case. |
| 2 | Duplicates | Drop | Entire Data | It will affect the accuracy |

1. **Feature Selection:**

| **S No** | **Selected Feature Name** | **Reason** | **Test Performed** |
| --- | --- | --- | --- |
| 1 | Text | NLP text analysis | - |
| 2 | Label | Output variable | - |

**Phase 3: Model Building:**

| **S No** | **Type of Problem** | **Approach** | **Algorithm Name** |
| --- | --- | --- | --- |
| 1 | Classification | Probability based | Naive Bayes |
| 2 | Classification | Boundary Based | Logistic Regression |
| 3 | Classification | Ensemble | Random Forest |
| 4 | Classification | Rules Based | Decision Tree |
| 5 | Classification | Ensemble | xgBoost |

**Phase 4: Model Evaluation:**

| **S No** | **Algorithm Name** | **Metric Score** | **Hyper Parameters** |
| --- | --- | --- | --- |
| 1 | Naive Bayes | 96 | Alpha=1 |
| 2 | Logistic Regression | 99 | c =0.1,class\_weight=  Balanced |
| 3 | Random Forest | 98 | Alpha = 1 |
| 4 | Decision Tree | 97 | max\_depth=5 |
| 5 | xgboost | 98 | learning\_rate=0.3 |

**Phase 5: Model Deployment:  
  
Deployment Platform:**  Streamlit

**Link/URL:** provide the link