**Project Document**



**Student Name:** Thota Chandra Sekhar

**Batch:** AI Elite 16(251)



**Project Name:** Campus Placement Status Prediction using Gaussian Naïve Bayes

**Domain:** Education

**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:** Develop a predictive model to determine the likelihood of successful campus placements for students based on their academic performance.

**Constraints:**   
  
**Success Criteria:**

1. ML success criteria: with average 85% accuracy(estimation)
2. Business Success criteria: -
3. Economy Success criteria: -
4. **Data Understanding:**

|  |  |  |
| --- | --- | --- |
| **S No** | **Feature Name** | **Data Type** |
| 1 | Sl\_no | Discrete |
| 2 | Gender | Categorical |
| 3 | Ssc\_p | Continuous |
| 4 | Ssc\_b | Categorical |
| 5 | Hsc\_p | Continuous |
| 6 | Hsc\_b | categorical |
| 7 | Hsc\_s | Categorical |
| 8 | Degree\_p | Continuous |
| 9 | Degree\_t | Categorical |
| 10 | Workex | Categorical |
| 11 | Etest\_p | continuous |
| 12 | Specialization | Categorical |
| 13 | Mba\_p | Continuous |
| 14 | Status | Categorical |
| 15 | Salary | Continuous |

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

|  |  |  |  |
| --- | --- | --- | --- |
| **S No** | **Type** | **Feature Names** | **Observation** |
| 1 | Missing Values | Salary | 67 |
| 2 | Duplicates |  | - |
| 3 | Outliers | Salary | - |
| 4 | Distributions | Salary | Right skewed |
| 5 | Noisy data | - | - |

**b) Data Cleaning/wrangling:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S no** | **Type of Cleaning** | **Technique** | **Feature Name** | **Reason** |
| 1 | Missing value | - | - | - |
| 2 | Encoding | Ordinal | Workex, specialization | To get numbers |
| 3 | Scaling | Robust Scaling | Ssc\_p,hsc\_p,degree\_p | To keep all values in certain range |

1. **Feature Selection:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S No** | **Removed Feature Name** | **Reason** | **Test Performed** |
| 1 | Etest\_p,mba\_p,salary | - | F\_classification |
| 2 | Gender,ssc\_b,hsc\_b,shc\_s,degree\_t | - | Chi2 |
|  |  |  |  |
|  |  |  |  |

**Phase 3: Model Building:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S No** | **Type of Problem** | **Approach** | **Algorithm Name** |
| 1 | Classification | Probability | Mixed Naïve Bayes |
| 2 | Classification | Distance Based | KNN |
|  |  |  |  |

**Phase 4: Model Evaluation:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S No** | **Algorithm Name** | **Metric Score** | **Hyper Parameters** |
| 1 | KNN | 81 | K =3 |
| 2 | Naïve Bayes | 81 | Alpha = 0.5 |
|  |  |  |  |
|  |  |  |  |

**Phase 5: Model Deployment:  
  
Deployment Platform:**  Stream lit

**Link/URL:** provide the link