Veridia Resume Classifier – Internship Task Report

Submitted by: **Chaitanya Bharat Dusane**

Date: October 20, 2025

Task Option: **Option B – End-to-End Solution**

1. Tech Stack / Frameworks Used

Category	Tools & Libraries
Programming Language	Python 3.10+
Data Processing	pandas, NumPy
NLP	NLTK, BeautifulSoup
Machine Learning	scikit-learn (Logistic Regression, TF-IDF Vectorizer)
PDF Handling	PyPDF2
Model Persistence	joblib
Web Framework	Flask
Frontend	HTML5, CSS3, JavaScript (tsParticles)
Environment Management	virtualenv
Version Control	Git, GitHub

2. Approach Followed

Step 1: Dataset from Kaggle containing 2,484 resumes across 25+ job categories.
Step 2: Cleaned and preprocessed resume text (HTML removed, stopwords eliminated).
Step 3: Extracted features using TF-IDF Vectorization (max_features=10,000, ngram_range=(1,3)).
Step 4: Trained Logistic Regression model on 80-20 stratified split.
Step 5: Developed Flask web app with drag-and-drop upload, dark theme UI, and confidence indicator.

3. Results & Accuracy Metrics

Overall Metrics:
Accuracy: 92.67%
Macro F1-Score: 92.1%
Weighted F1-Score: 92.7%

Top Category Sample Report:
ACCOUNTANT (Precision: 0.96, Recall: 0.94, F1: 0.95)
ADVOCATE (Precision: 0.98, Recall: 0.97, F1: 0.97)
AGRICULTURE (Precision: 0.93, Recall: 0.91, F1: 0.92)

ARTS (Precision: 0.95, Recall: 0.96, F1: 0.95)
AUTOMOBILE (Precision: 0.97, Recall: 0.95, F1: 0.96)

4. Screenshots / Outputs of Model Performance

Below is the screenshot of the output interface showing classification results.