Project Report: Campus Placement

Name: Kiranpreet Kaur
ID: c0925339

1. **Dataset Overview**

- **Source**: Kaggle Campus Recruitment Data
- **Samples**: 215 student records
- **Target**: Placement Status (Placed/Not Placed)
- **Features**: 12 features including academic scores, work experience, and specialization
- **Class Distribution**: 148 Placed (68.8%), 67 Not Placed (31.2%)

2. Key Preprocessing Steps

- Missing values in MBA percentage handled with median imputation
- Categorical variables encoded using One-Hot Encoding
- Numerical features standardized using Z-score normalization
- Stratified train-test split (70-30) to preserve class distribution

3. Model Selection & Tuning

- **Logistic Regression**: Best with C=10 and liblinear solver
- **Random Forest**: Optimal with 200 trees and max_depth=10
- **SVM**: Best performance with RBF kernel and C=10
- All models used class weighting to handle slight class imbalance

4. Performance Evaluation

	Accuracy	Precision	Recall	F1
Logistic Regression	0.830769	0.904762	0.844444	0.873563
Random Forest	0.846154	0.843137	0.955556	0.895833
SVM	0.800000	0.880952	0.822222	0.850575
Voting	0.800000	0.847826	0.866667	0.857143

5. Key Insights

❖ **Random Forest Emerges as Top Performer:**

- Highest F1-score (0.90) despite slightly lower precision

- Best recall (96%) indicating strong identification of placed students

❖ **Voting Classifier Performance:**

- Third position in overall performance
- Balanced precision (85%) and recall (87%)
- Shows 4% improvement over worst individual model (SVM)

❖ **Model Characteristics:**

- Clear precision-recall tradeoff visible:
- Logistic Regression: High precision (90%) but lower recall
- Random Forest: High recall (96%) with moderate precision
- No single model dominates all metrics

6. Conclusion

- Recommended model: Random Forest (Optimal F1-score: 90%) for the strongest balance between placement detection and false positives.
- Practical Implications:
  - ✓ Use random forest for screening of initial candidates
  - ✓ Finish off with logistic regression after shortlisting because of the improved precision
  - ✓ Weighting can improve the performance of voting classifier.
- Actionable Insights: Target educational qualifications (the top ranked features of random forest) for placement activities.