Placement Chance Prediction Using Machine Learning

📑 Problem Statement

To develop a machine learning-based system that predicts the **likelihood of student placement** based on their academic performance, skills, and other relevant features. The goal is to assist students in understanding their preparedness for campus placements and identify areas of improvement.

🔍 Research Objectives

1. Build a regression model to predict the **placement chance** of a student.

2. Evaluate the model performance and continuously improve it.

3. Use both linear and ensemble models to compare results.

4. Deploy an interactive Streamlit-based web app for easy access and use.

🧠 Hypothesis

- **H_0 (Null Hypothesis):** Student features (CGPA, Programming Skill, Communication Skill, etc.) have **no significant effect** on placement chance.

- **H_1 (Alternate Hypothesis):** These features **significantly affect** placement chances.

📙 Data Overview

Initially, a **small synthetic dataset** was created manually with minimal data points including the following features:

- CGPA

- ProgrammingSkill

- CommunicationSkill

- AptitudeScore

- MockTestScore

- ExtraCurricular

- InternshipCount

- ProjectCount

Target variable:

- **PlacementChance** (ranged from 0 to 1)

📊 **Phase 1: Initial Experiment with Linear Regression**

Model: LinearRegression()

Results:

- **R² Score:** -1.1956

- **MSE:** 0.1043

- **RMSE:** 0.3229

- **MAE:** 0.2573

❌ **Interpretation:**

- The model **performed worse than baseline average prediction**.

- R² < 0 suggests the model lacks the predictive capability.

- The dataset was **too small** and lacked diversity.

📈 **Phase 2: Improved Dataset + Better Model**

Actions Taken:

- Expanded dataset with **more records**.

- Cleaned data and scaled features.

- Switched to **RandomForestRegressor** for better performance.

Results (Random Forest Regressor):

- **R² Score:** 0.6701

- **MSE:** 0.0032

- **RMSE:** 0.0565

- **MAE:** 0.0444

✅ **Interpretation:**

- Much **better fit** and **reduced error**.

- Predictions closely matched actual placement chances.

- The model is **ready for deployment**.

❄️ **Streamlit App Highlights**

- User inputs: sliders and dropdowns for academic and skill scores.

- Real-time prediction of placement chance.

- Visual and animated feedback:

- o ⭐ High Chance: Celebratory animation

- o 😊 Fair Chance: Motivational animation

- o ⚠️ Low Chance: Encouragement to improve skills

🔬 Final Findings

- **Model performance improved significantly** with increased dataset size and a better model (Random Forest).

- **Academic scores and skills have a strong influence** on placement chances.

- Linear regression is too simplistic for this problem.

- Ensemble models like RandomForest offer robustness and better generalization.

📝 Recommendations

- Expand the dataset with real-world placement records (if available).

- Include features like domain knowledge, resume score, soft skills.

- Apply **hyperparameter tuning** for further boosting accuracy.

- Deploy the Streamlit app with animations on a cloud platform.

- Use **feature importance** scores to guide student improvement areas.

📖 Summary Table

Phase	Model	R² Score	MAE	RMSE	Dataset Size	Comments
1	Linear Regression	-1.19	0.2573	0.3229	Small	Poor performance due to low data
2	RandomForest Regressor	0.6701	0.0444	0.0565	Expanded	Great improvement and reliable model

---

Prepared by: Vishwanath Kulal

Tools used: Python, Scikit-Learn, Streamlit, Pandas, Seaborn

Models: Linear Regression, Random Forest, XGBoost