# **Scripts to train models and make predictions:**

describes the purpose of the scripts. These scripts are responsible for training machine learning models using available data and then using those trained models to make predictions on new or unseen data.

**Trained Model:**

A trained model can predict the employee performance based on factors as inputs. This will be used to hire employees as well as drive insights on their efficiency.

**Predictive Model:**

This model will be a useful tool for making decisions and forecasts in real time.

**[**The following models are trained and stored in separate files and their predictions also been stored in separate files. Refer the “**models**” folder in “**src**” parent directory**]**

# **Model** Performance **Evaluation:**

**Decision Tree:**

**Accuracy score:** 91.2%

**Strengths:** Capable of complex data, capturing non-linear relationships in the data.

**Weakness**: Slightly overfitting, difficult to handle high dimensional data.

**Considering:** Additional processing power might be needed.

**Tuned Decision Tree:**

**Accuracy score:** 90.8%

**Strengths**: Better generalization, reduces the risk of overfitting.

**Weakness**: careful tuning is required, computational resource exhaustive.

**Considering:** Additional processing power might be needed.

**Random Forest:**

**Accuracy score:** 93.7%

**Strengths:** Best model compared to others, combines multiple decision trees thus reduces overfitting; can handle high dimensional data.

**Weakness:** More computational resource is required; difficulty in tuning due to large number of hyperparameters.

**Considering:** Additional processing power might be needed.

**Tuned Random Forest:**

**Accuracy score:** 93.3%

**Strengths:** performance almost equivalent to base model here; better generalization.

**Weakness:** Increased experimentation with tuning.

**Considering:** Additional processing power might be needed.

**Support Vector Machine:**

**Accuracy score:** 76.6%

**Strengths:** Capable of handling high-dimensional data, versatile and memory efficient.

**Weakness:** computationalcomplexity; sensitive to kernel choice and it’s parameters.

**Considering:** Additional processing power might be needed.

**Tuned Support Vector Machine:**

**Accuracy score:** 86.6%

**Strengths:** Improved performance compared to the base model; ablility to handle complex relationships in the data.

**Weakness:** Needs more computational resource.

**Considering:** Additional processing power might be needed.

**Extreme Gradient** **boosting**

**Accuracy score:** 93.3%

**Strengths:** High performance in predictive accuracy; strong against overfitting.

**Considering:** Additional processing power might be needed.



