CROP PREDICTION SYSTEM

Using Machine Learning

WELCOME

In Our Group Project

Meet The Group

NAMRATA
ASATI

AKASH
MERUGA

SAI KIRAN
SADA

Introduction

Machine learning is a valuable decision-making tool for predicting agricultural yields and deciding the type of crops to sow and things to do during the crop growing season. In order to aid crop prediction studies, several machine learning methods have been used.

Machine learning techniques are utilized in various sectors, from evaluating customer behavior in supermarkets to predicting customer phone usage. For some years, agriculture has been using machine learning techniques. Crop prediction is one of agriculture's complex challenges, and several models have been developed and proven so far. Because crop production is affected by many factors such as atmospheric conditions, type of fertilizer, soil, and seed, this challenge necessitates using several datasets. This implies that predicting agricultural productivity is not a simple process; rather, it entails a series of complicated procedures. Crop yield prediction methods can now reasonably approximate the actual yield, although more excellent yield prediction performance is still desired.

Table Of Contents

Crop prediction system Using

Title Of The Chart Page

01	02	03	04
Data Preception and pre processing	Feature Engineering	Model Training	Crop Yeild Prediction

Using SVM

Data Collection and Preprocessing: Gathering relevant agricultural data, including historical crop yields, soil characteristics, weather data, and more. This data is preprocessed to ensure it's in a suitable format for SVM analysis, which may include normalization, feature selection, and cleaning.

Feature Engineering: Identifying and selecting important features or variables from the dataset that significantly affect crop yields, such as temperature, humidity, soil pH, and crop type. SVM models rely on these features to make predictions.

Model Training: Using the preprocessed data, SVM models are trained to find the optimal hyperplane that best separates different crop yield outcomes or classes. SVMs aim to maximize the margin between these classes, helping to predict crop performance accurately.

Crop Yield Prediction: Once the SVM model is trained, it can be used to predict crop yields for future seasons based on input data. Farmers and agricultural stakeholders can then make informed decisions regarding planting, [illegible]n, and [illegible]. [illegible] resource allocation to optimize their crop production and overall agricultural

A kernel in SVM is a mathematical function that transforms the input data into a higher-dimensional space where it is easier to find a separating hyperplane between different classes. A kernel can help SVM deal with nonlinear and complex data that cannot be separated by a linear boundary in the original space.

There are different types of kernels in SVM, such as linear, polynomial, radial basis function (RBF), sigmoid, and custom kernels.

SVM (linear) :

A Linear Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification tasks. It aims to find the optimal linear boundary that maximizes the margin between different classes, making it a powerful tool for separating data points into distinct categories with high accuracy and robustness.

SVM (poly) :

The main point of a polynomial support vector machine (Poly SVM) is to classify data points in a non-linearly separable dataset by transforming them into a higher-dimensional space using polynomial functions. This allows for the creation of non-linear decision boundaries, enabling accurate classification of complex data patterns.

SVM (rbf) :

Support Vector Machines (SVM) with Radial Basis Function (RBF) kernel are a powerful classification technique. RBF SVMs are effective for non-linear data separation, as they map data into higher dimensions, finding optimal decision boundaries. They excel in various applications, offering robust performance and strong generalization capabilities.

Random Forest

Data Collection and Preprocessing: The first step in using Random Forest for crop prediction is collecting relevant data. This includes historical crop yield data, climate information, soil characteristics, and other relevant factors. This data is then preprocessed to clean, normalize, and transform it into a suitable format for machine learning.

Feature Selection: Selecting the most important features or variables for prediction is crucial. Random Forest can automatically rank the importance of features, helping identify which factors have the most influence on crop yield. This assists in optimizing input variables for the model.

Model Training: After data preparation and feature selection, the Random Forest model is trained on the historical dataset. The algorithm builds an ensemble of decision trees, which collectively make predictions based on the input features. Random Forest is known for its robustness, as it reduces overfitting and handles noisy data well.

Crop Yield Prediction: Once the Random Forest model is trained, it can be used to predict crop yields for future planting seasons. Farmers can input current environmental conditions (e.g., weather, soil data) into the model to receive predictions on which crops are likely to perform well, helping them make informed decisions about planting and resource allocation.

Random Forest offers the advantage of being interpretable, versatile, and capable of handling both classification and regression tasks.

Multi-Layered Perception

1. **Data-Driven Forecasting:** MLPs are employed to analyze extensive datasets, including information on weather, soil conditions, historical crop yields, and various other factors. By training on this data, MLPs can recognize patterns and relationships, allowing them to make predictions about future crop yields and conditions.
2. **Feature Extraction:** MLPs can extract relevant features from the input data, highlighting the most important factors influencing crop growth. This helps in identifying the key variables that affect crop production, such as temperature, precipitation, and soil nutrients.
3. **Customization for Local Conditions:** MLP models can be tailored to specific regions or farms, accounting for localized climate and soil characteristics. This adaptability ensures that predictions are accurate and relevant to the particular context in which they are applied.
4. **Improved Decision Making:** By providing accurate crop yield predictions, MLP-based systems empower farmers to make informed decisions regarding planting schedules, irrigation, and resource allocation. This ultimately enhances agricultural productivity and resource efficiency.

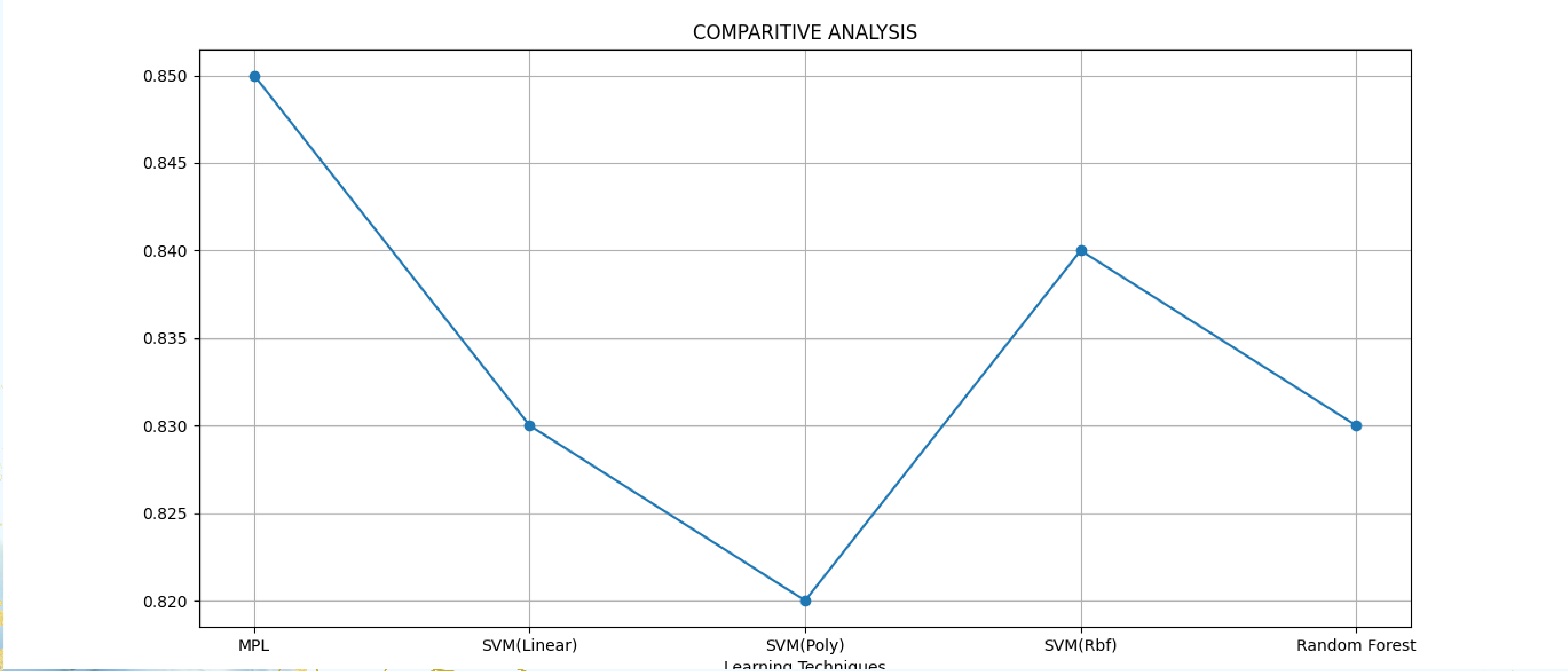
Machine learning (ML) is the process of creating systems that can learn from data and improve their performance without being explicitly programmed. ML models are algorithms that can find patterns, make predictions, or perform tasks based on the data they are trained on, you need to follow some general steps that are part of the ML lifecycle:

1. Define The Problem and goal
2. Collect and Prepare the Data
3. Choose And Train your your ML Model
4. Evaluate Sand Improve your ML Model
5. Display and Monitering your ML

This table represents the accuracy achieved on our dataset while working with five different learning algorithm out of a total of seven features.

LEARNING	ACCCURACY
MPL	0.85
SVM (LINEAR)	0.8345
SVM (poly)	0.8298
SVM (rbf)	0.8408
Random Forest	0.8378

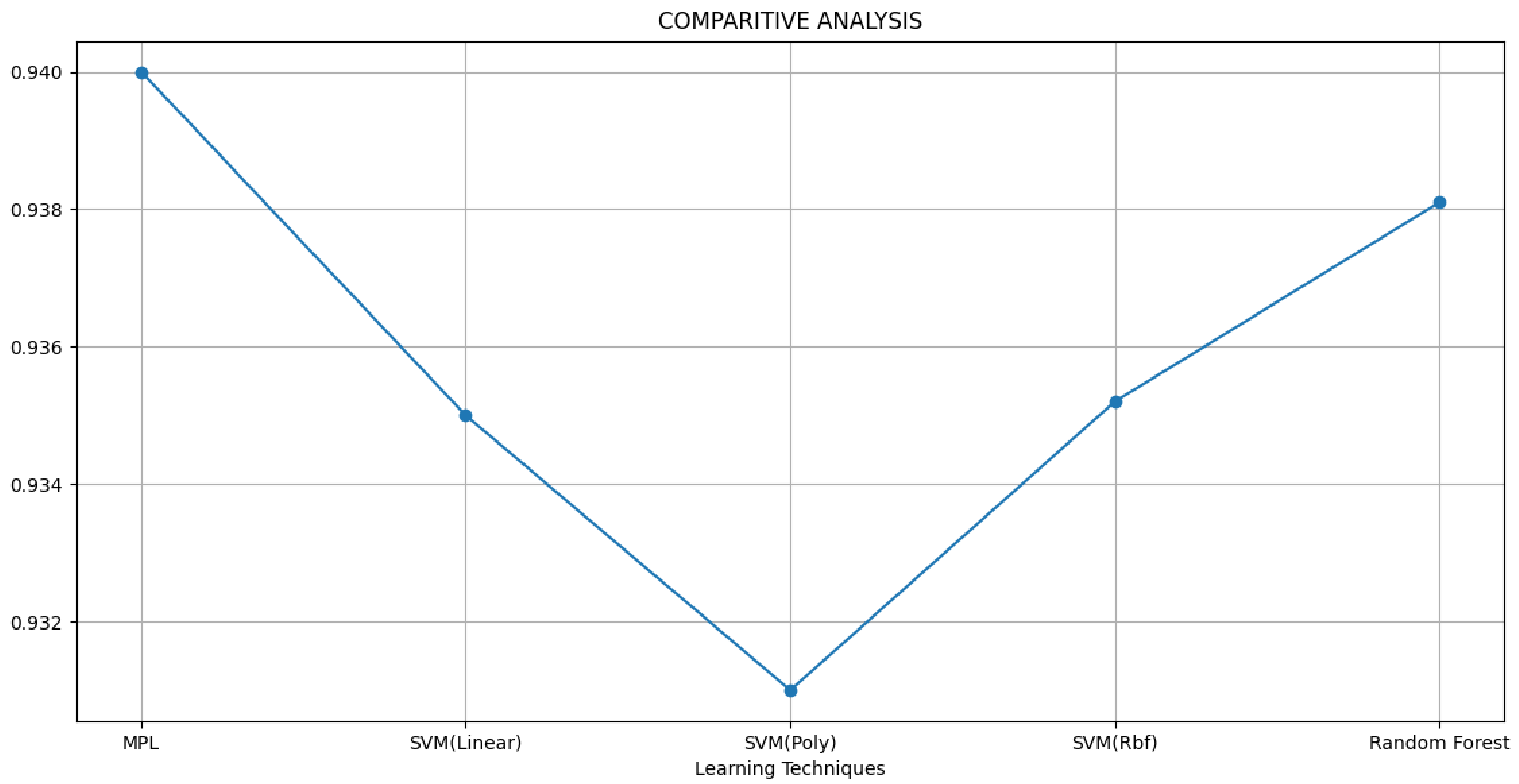
Graphical Representation of above mentioned table



This table represents the accuracy achieved on our dataset while working with Six different learning algorithm out of a total of Seven features.

LEARNING	ACCCURACY
MPL	0.94
SVM (LINEAR)	0.9345
SVM (poly)	0.9311
SVM (rbf)	0.9352
Random Forest	0.9381

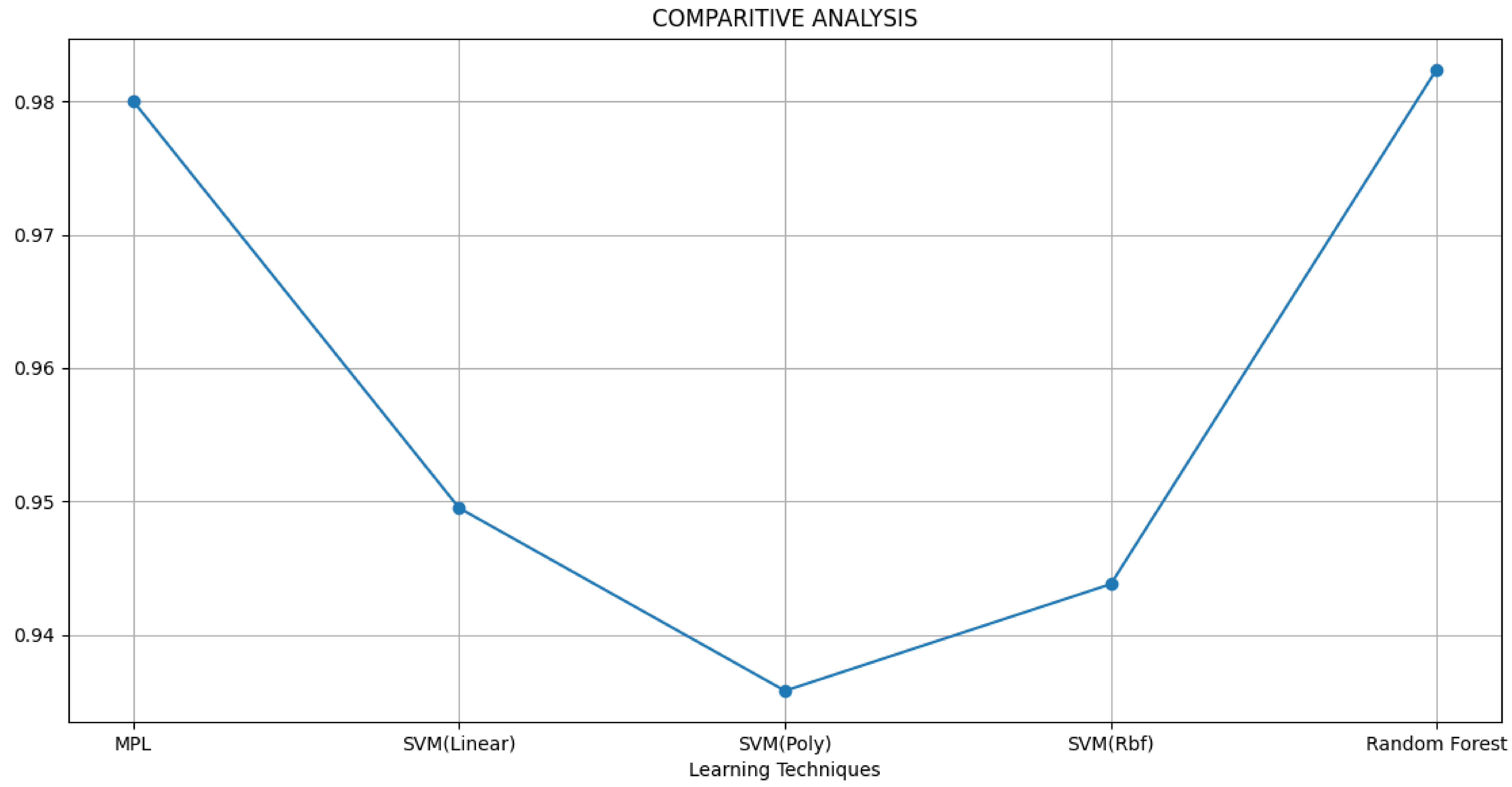
Graphical Representation of above mentioned table



This table represents the accuracy achieved on our dataset while working with seven learning algorithm.

LEARNING	ACCCURACY
MPL	0.98
SVM (LINEAR)	0.94
SVM (poly)	0.9358
SVM (rbf)	0.9438
Random Forest	0.9824

Graphical Representation of above mentioned table



Conclusion:

In my machine learning project, I applied different algorithms to the same data set and compared their performance. I found that random forest regression and multilayer perceptron (MLP) achieved the highest accuracy among all the methods. On the other hand, support vector machine (SVM) with polynomial kernel had the lowest accuracy. I also observed that reducing the number of features resulted in lower accuracy for all the algorithms.

- **Random Forest:** Random Forest can handle high-dimensional data and noisy data well. It also has a built-in mechanism to prevent overfitting by randomly selecting a subset of features at each split

Random Forest is a type of ensemble learning method that combines multiple decision trees to make predictions. It is known for its ability to handle high-dimensional data and noisy data, as well as its resistance to overfitting

MLP: MLP can learn complex patterns in data and can handle different types of data well. It is also flexible in terms of the number of hidden layers and the number of neuron

MLP, or Multi-Layer Perceptron, is a type of neural network that is commonly used for classification tasks. It is known for its ability to learn complex patterns in data and its flexibility in handling different types of data

It's important to note that the effect of removing a feature on the accuracy of a model can vary depending on the specific dataset and the algorithm used. Some algorithms are more sensitive to changes in the input data than others, and some datasets may contain redundant or irrelevant features that do not contribute much to the accuracy of the model.

In our case, it's possible that MLP and Random Forest were able to perform well on your dataset because they were able to learn from the information contained in all of the features. When you removed one of the features, this caused a decrease in accuracy because the models had less information to work with.

THANK YOU