# UNIT-04: ﻿Installation and development environment overview

﻿**Rapid Miner** provides an environment for machine learning and data mining processes.

﻿**Facilities of RapidMiner**

1. Rapid Miner provides its collection of datasets, but it also provides options to set up a database in the cloud for storing large amounts of data. You can store and load the data from Hadoop, Cloud, RDBMS, NoSQL, etc. Apart from this, you can load your CSV data very easily and start using it as well.

﻿2. The standard implementation of procedures like data cleaning, visualization, pre-processing can be done with drag and drop options without having to write even a single line of code.

3. Rapid Miner provides a wide range of machine learning algorithms in classification, clustering, and regression as well. You can also train optimal deep learning algorithms like Gradient Boost, XGBoost, etc. Not only this, but the tool also provides the ability to perform pruning and tuning.

4. Finally, to bind everything together, you can easily deploy your machine learning models to the web or mobiles through this platform. You just need to create user interfaces to collect realtime data and run it on the trained model to serve a task.

**Rapid Miner Products**

﻿**RapidMiner Studio**

• With RapidMiner Studio, one can access, load, and analyze both traditional structured data and unstructured data like text, images, and media.

• It can also extract information from these types of data and transform unstructured data into structured. **RapidMiner Auto Model**

• Auto Model is an advanced version of RapidMiner Studio that increments the process of building and validating data models.

• Majorly three kinds of problems can be resolved with Auto Model namely prediction, clustering, and outliers.

**RapidMiner Turbo Prep**

• Data preparation is time-consuming, and RapidMiner Turbo Prep is designed to make the preparation of data much easier.

• It provides a user interface where your data is always visible front and center, where you can make changes step-by-step and instantly see the results, with a wide range of supporting functions to prepare the data for model-building or presentation.

﻿**WEKA** stands for Waikato Environment for Knowledge Analysis. It provides a lot of tools for data preprocessing, classification, clustering, regression analysis, association rule creation, feature extraction, and data visualization. It is a powerful tool that supports the development of new algorithms in machine learning.

A picture containing text, screenshot, diagram, line

Description automatically generated

﻿First, you will start with the raw data collected from the field. This data may contain several null values and irrelevant fields. You use the data preprocessing tools provided in WEKA to cleanse the data. Then, you would save the preprocessed data in your local store for applying ML algorithms. Next, depending on the kind of ML model that you are trying to develop you would select one of the options such as Classify, Cluster, or Associate. The Attributes Selection allows the automatic selection of features to create a reduced dataset. Note that under each category, WEKA provides the implementation of several algorithms. You would select an algorithm of your choice, set the desired parameters, and run it on the dataset. Then, WEKA would give you the statistical output of the model processing. It provides you a visualization tool to inspect the data.