Soft Computing

Dr. Hemraj S. L.

CSIT, SUAS

Cryptologist
Ethical Hacker
Compiler Designer
Theory of Machine
Information Security
DCIM
VA & Penetration Tester
Cloud Computing & Auditor
Fog and Edge Computing
Embedded Device Security
Blockchain Technology
IoT & Infrastructure Designer
Part-time Bug Hunter (HackerOne)

Assistant Professor and Still learning new bytes.

Course Plan

1. **Introduction to Soft Computing**
2. **Evolutionary Computing**
   - Genetic Algorithms (GAs)
   - Simulated Annealing (SA)
   - Ant Colony Optimization (ACO)
   - Particle Swam Optimization (PSO)
3. **Fuzzy Logic**
   - Fuzzy Set, Fuzz Logic, Fuzzy Algebra
   - Fuzzy Reasoning and Fuzzy Classification
4. **Artificial Neural Networks (ANNs)**
   - Different ANNs
   - Learning with ANNs
5. **Advanced Topics**
   - Mixed(Hybrid) Soft Computing
   - FL-GA, FL-ANN, GA-ANN, FL-GA-ANN
   - Hidden Markov Modeling (HMM)
   - Support Vector Machine (SVM)

Today's Topics

Introduction to Soft Computing

- Concept of computing
- Important characteristics of "Computing"
- **Soft** computing vs. "**Hard**" Computing
- Few examples of Soft computing applications
- Characteristics of Soft computing
- Hybrid computing

Soft Computing

Soft Computing could be a computing model evolved to resolve the non-linear issues that involve unsure, imprecise and approximate solutions of a tangle. These sorts of issues square measure thought of as real-life issues wherever the human-like intelligence is needed to resolve it.

Concept of Computing

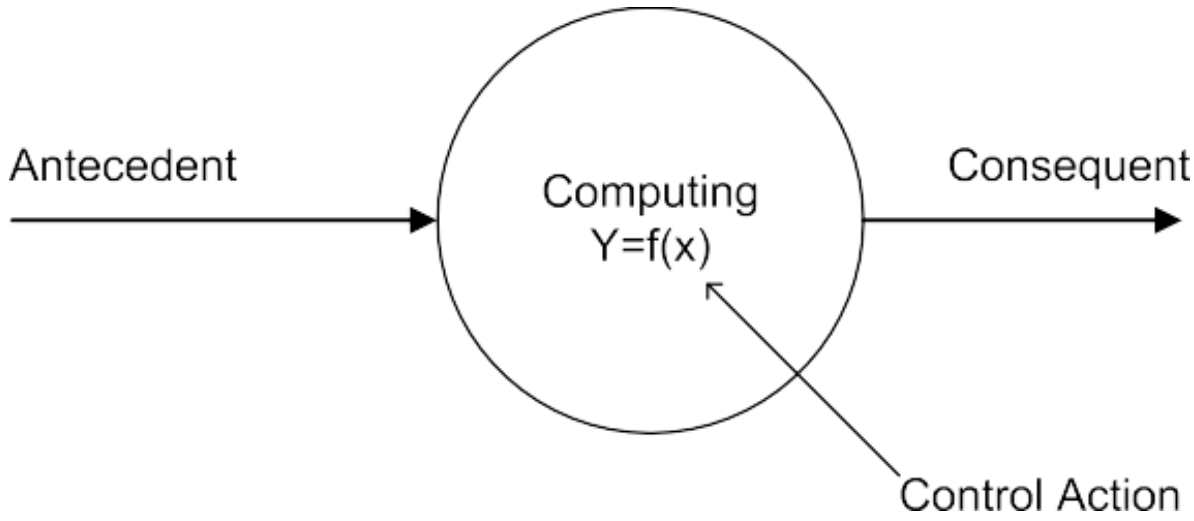


Figure :Basic of computing

$y = f(x)$, f is a mapping function

f is also called a formal method or an algorithm to solve a problem.

Important Characteristics

1. Should provide precise solution.
2. Control action should be unambiguous and accurate.
3. Suitable for problem, which is easy to model mathematically.

Hard Computing

Hard Computing is that the ancient approach employed in computing that accurately declared analytical model. the outcome of hard computing defines definite management actions employing a mathematical model or algorithmic rule. It deals with binary and crisp logic that need the precise input file consecutive.

Hard computing isn’t capable of finding the real world problem’s solution.

Hard Computing

In 1996, LA Zade (LAZ) introduced the term hard computing.

According to LAZ: We term a computing as ”Hard” computing, if

- **Precise result is guaranteed**
- **Control action is unambiguous**
- **Control action is formally defined (i.e. with mathematical model**

Example:

- Solving numerical problems (e.g. Roots of polynomials, Integration etc.)
- Searching and sorting techniques
- Solving ”Computational Geometry” problems (e.g. Shortest path in Graph theory, Finding closest pair of points etc.)

"Information is the oil of the 21st century, and Analytics is the combustion engine"

(Peter S., *Gartner*)

Problems in some other areas of applications

- Medical diagnosis
- Person identification / Computer vision
- Hand written character recognition
- Pattern recognition and Machine Intelligence MI
- Weather forecasting
- VLSI design
- Network optimization

Characteristics of Soft Computing

- It does not require any mathematical modeling of problem solving
- It may not yield the precise solution
- Algorithms are adaptive (i.e. it can adjust to the change of dynamic environment)
- Use some biological inspired methodologies such as genetics, evolution, Ant's behaviors, particles swarming, human nervous systems etc.

Hybrid Computing

It is a combination of the conventional hard computing and emerging soft computing

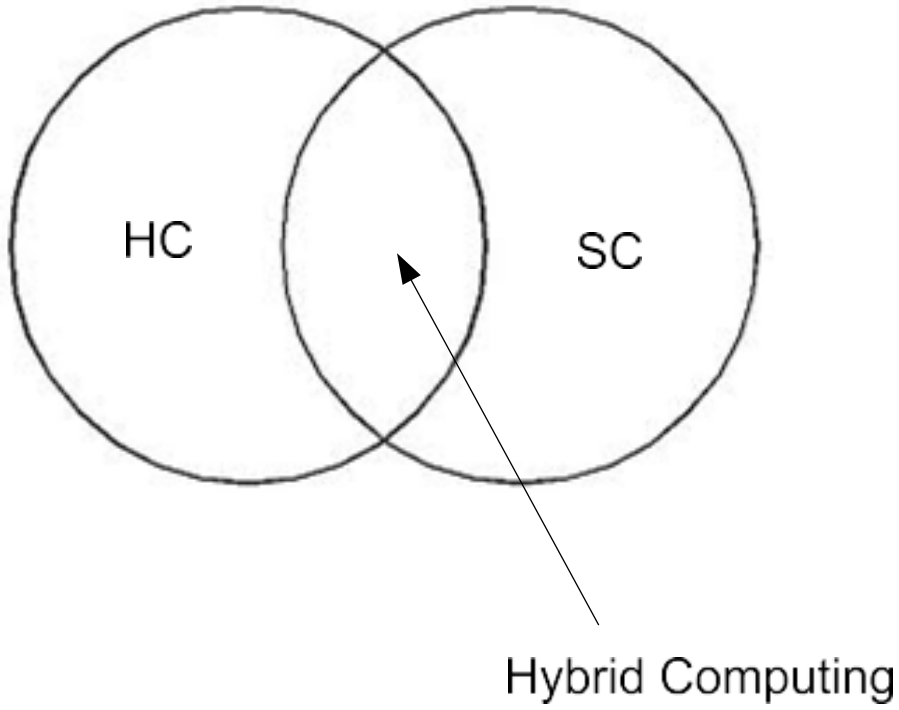


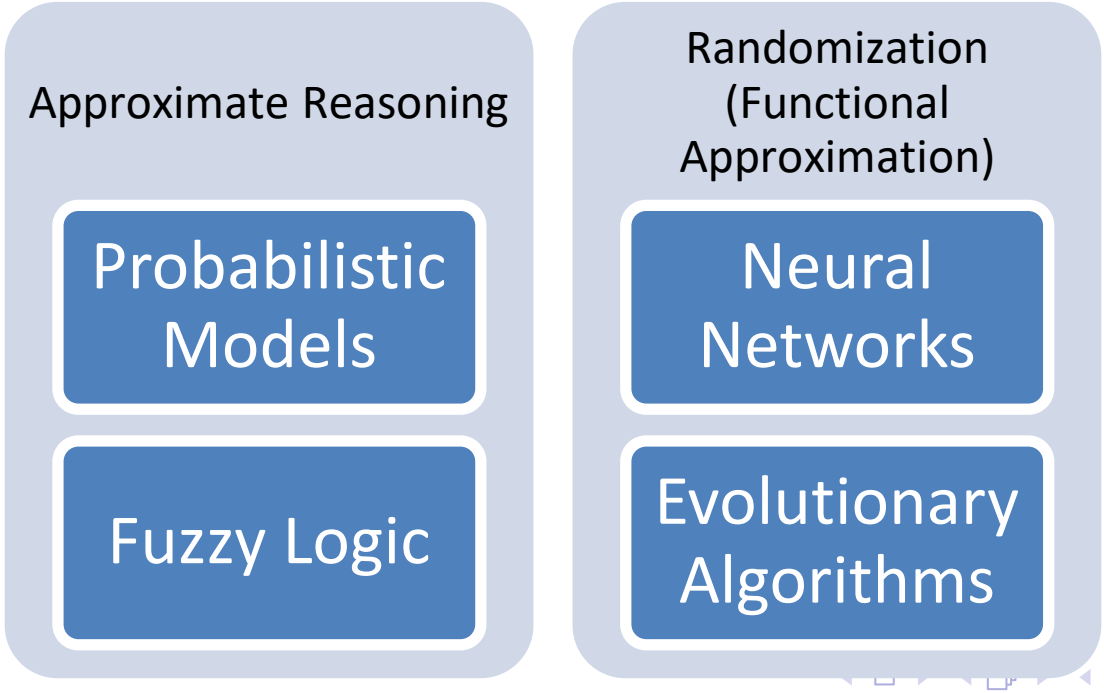
Figure :Concept of Hybrid Computing

Problems to ponder

- Hard computing (HC) vs. Soft computing (SC)
- Limitation(s) in HC and SC
- Examples of (**only**) Hard computing and (**only**) Soft computing
- Examples of Hybrid computing

Elements of Soft computing

Soft computing is viewed as a foundation component for an emerging field of conceptual intelligence. Fuzzy Logic (FL), Machine Learning (ML), Neural Network (NN), Probabilistic Reasoning (PR), and Evolutionary Computation (EC) are the supplements of soft computing. Also, these are techniques used by soft computing to resolve any complex problem.



Difference between SOFT computing and HARD computing

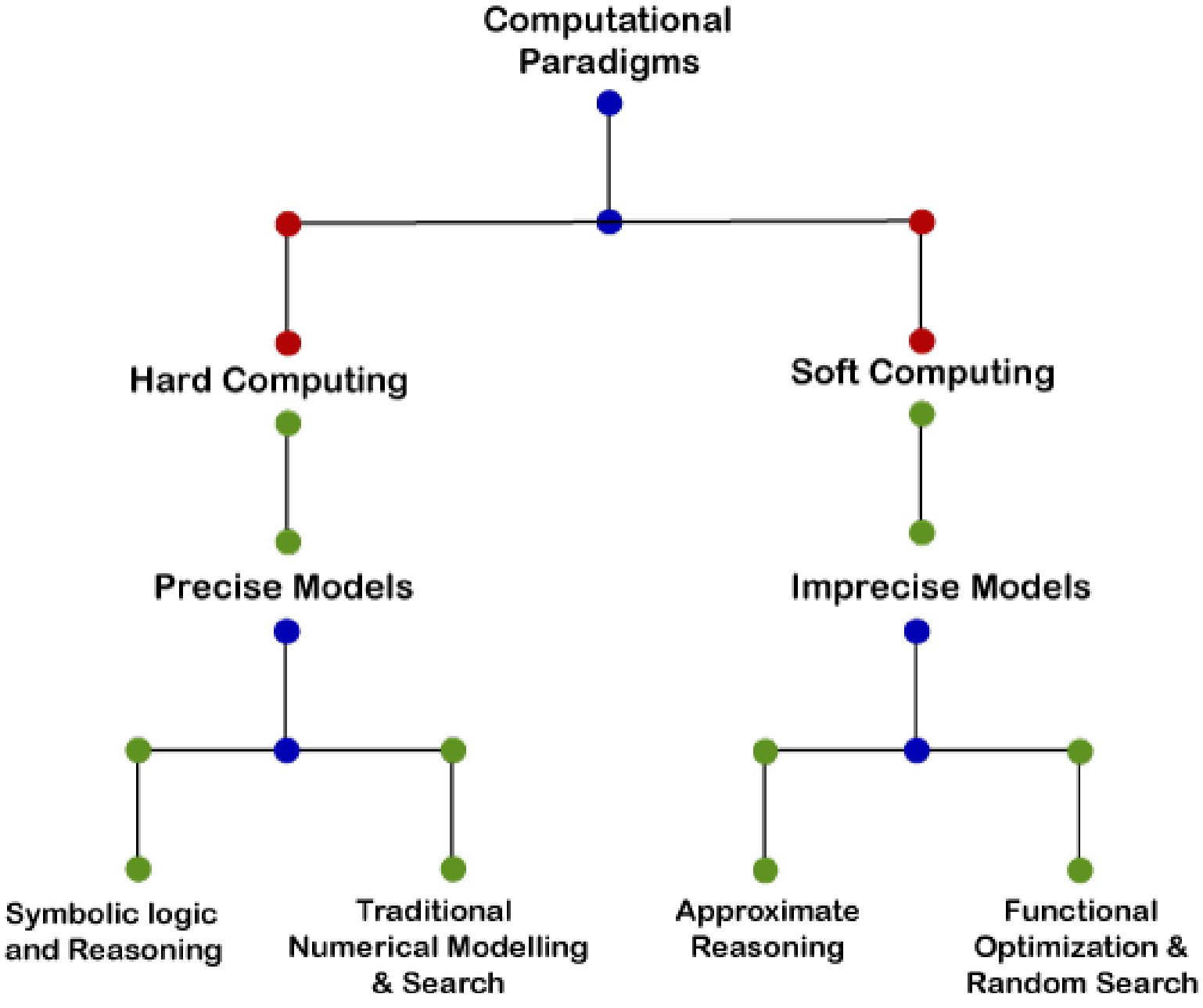
Soft Computing:

- It involves uncertainty, partial truth and approximation.
- relies on formal logic and probabilistic reasoning.
- has the features of approximation.
- Soft computing is stochastic in nature.
- works on ambiguous and noisy data.
- can perform parallel computations.
- produces approximate results.
- Soft computing incorporates randomness.
- uses multivalued logic.

Hard Computing:

- needs a exactly state analytic model.
- relies on binary logic and crisp system.
- has the features of precision.
- Hard computing is deterministic in nature.
- works on exact data.
- can performs sequential computations.
- produces precise results.
- requires programs to be written.
- uses binary valued logic.

Soft vs Hard Computing



Motivation: "Necessity is the Mother of Invention"

Data explosion problem

Automated data collection tools and mature database technology lead to tremendous amounts of data stored in databases, data warehouses and other information repositories

We are drowning in data, but starving for knowledge!

Solution: Data Analytics

Data Analytics and on-line analytical processing

Extraction of interesting knowledge (rules, regularities, patterns, constraints) from data in large databases

What Is Data Analytics?

Data Analytics (knowledge discovery in databases):

Extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from data in large databases

Alternative names and their "inside stories":

Data Analytics: a misnomer?

Knowledge discovery(mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, etc.

What is not data analytics?

(Deductive) query processing.

Expert systems or small machine learning/statistical programs

7 Data Analytics Steps

1. Data cleaning – remove noise and inconsistent data
2. Data integration – combine multiple sources
3. Data selection – retrieve from the database data relevant to the analysis task
4. Data transformation – data are transformed or consolidated into forms appropriate for mining (e.g. performing summary or aggregation operations)
5. Data Analytics – intelligent methods are applied to extract data patterns
6. Pattern evaluation – identify truly interesting patterns representing knowledge based on some interestingness measures
7. Knowledge presentation – present mined knowledge to the user

Why Data Analytics? — Potential Applications

- Database analysis and decision support
  - Market analysis and management
    - target marketing, customer relation management, market basket analysis, cross selling, market segmentation
  - Risk analysis and management
    - Forecasting, customer retention, improved underwriting, quality control, competitive analysis
  - Fraud detection and management
- Other Applications
  - Text mining (news group, email, documents) and Web analysis.
  - Intelligent query answering

Data Analytics Tools

Data Analytics is the set of techniques that utilize specific algorithms, statistical analysis, artificial intelligence, and database systems to analyze data from different dimensions and perspectives.

Data Analytics tools have the objective of discovering patterns/trends/groupings among large sets of data and transforming data into more refined information.

Data Analytics Tools

It is a framework, such as Rstudio or Tableau that allows you to perform different types of data mining analysis.

We can perform various algorithms such as clustering or classification on your data set and visualize the results itself. It is a framework that provides us better insights for our data and the phenomenon that data represent. Such a framework is called a data mining tool.

The Market for Data Mining tool is shining: as per the latest report from ReportLinker noted that the market would top **$1 billion** in sales by 2023, up from **$ 591 million** in 2018.

Data Analytics Tools & more

Orange Data Mining

Orange is a perfect machine learning and data mining software suite. It supports the visualization and is a software-based on components written in Python computing language and developed at the bioinformatics laboratory at the faculty of computer and information science, Ljubljana University, Slovenia.

As it is a software-based on components, the components of Orange are called "widgets." These widgets range from preprocessing and data visualization to the assessment of algorithms and predictive modeling.

Widgets deliver significant functionalities such as:

- Displaying data table and allowing to select features
- Data reading
- Training predictors and comparison of learning algorithms
- Data element visualization, etc.

Orange Data Mining

Python scripts can keep running in a terminal window, an integrated environment like PyCharmand PythonWin, pr shells like iPython. Orange comprises of canvas interface onto which the user places widgets and creates a data analysis workflow. The widget proposes fundamental operations, For example, reading the data, showing a data table, selecting features, training predictors, comparing learning algorithms, visualizing data elements, etc.

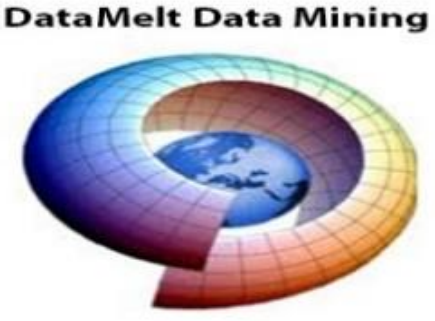
Orange can read documents in native and other data formats. Orange is dedicated to machine learning techniques for classification or supervised data mining. There are two types of objects used in classification: learner and classifiers.

SAS - Statistical Analysis System

It is a product of the SAS Institute created for analytics and data management. SAS can mine data, change it, manage information from various sources, and analyze statistics.

SAS data miner allows users to analyze big data and provide accurate insight for timely decision-making purposes. SAS has distributed memory processing architecture that is highly scalable. It is suitable for data mining, optimization, and text mining purposes.

DataMelt Data



DataMelt is a computation and visualization environment which offers an interactive structure for data analysis and visualization.

DMelt can be used for the analysis of the large volume of data, data mining, and statistical analysis. It is extensively used in natural sciences, financial markets, and engineering.

DMelt is a multi-platform utility written in JAVA. It can run on any operating system which is compatible with JVM. It consists of some libraries include:

Scientific libraries: Scientific libraries are used for drawing the 2D/3D plots.
Mathematical libraries: Mathematical libraries are used for random number generation, algorithms, curve fitting, etc.

Togaware: Rattle

Rattle is a data mining tool based on GUI. It uses the R stats programming language. Rattle exposes the statistical power of R by offering significant data mining features. While rattle has a comprehensive and well-developed user interface, It has an integrated log code tab that produces duplicate code for any GUI operation.

The data set produced by Rattle can be viewed and edited. Rattle gives the other facility to review the code, use it for many purposes, and extend the code without any restriction.

Rapid Miner

Rapid Miner is one of the most popular predictive analysis systems created by the company with the same name as the Rapid Miner. It is written in JAVA programming language. It offers an integrated environment for text mining, deep learning, machine learning, and predictive analysis.

The instrument can be used for a wide range of applications, including company applications, commercial applications, research, education, training, application development, machine learning.

It support cloud based analysis.

IBM SPSS MODELER

IBM SPSS Modeller is a visual data science and machine learning solution, helping in shortening the time to value by speeding up operational tasks for data scientists. IBM SPSS Modeler will have you covered from drag and drop data exploration to machine learning.

The software is used in leading enterprises for data preparation, discovery, predictive analytics, model management and deployment. The tool helps organizations to tap into their data assets and applications easily. One of the advantages of proprietary software is its ability to meet robust governance and security requirements of an organization at the enterprise level

KNIME

Konstanz Information Miner is an open-source data analysis platform, helping you with build, deployment and scale in no time. The tool aims to help make predictive intelligence accessible to inexperience users. It aims to make the process easy by it is a step by step guide based GUI tools.

The product markets itself as an End to End Data Science product, that helps create and production data science using its single easy and intuitive environment.

Python

Python is a freely available and open-source language that is known to have a quick learning curve. Combined with is the ability as a general-purpose language and it is a large library of packages that help build a system for creating data models from the scratch, Python makes for a great tool for organizations who want the software they use to be custom built to their specifications.

One of the features Python is known for in this field is powerful on the fly visualization features it offers.

Kaggle

The largest community of data scientists and machine learning professionals. Kaggle although started as a platform for machine learning competitions, is now extending its footprint into the public cloud-based data science platform arena. Kaggle now offers code and data that you need for your data science implementations.

There are over 50k public datasets and 400k public notebooks that you can use to ramp up your data mining efforts. The huge online community that Kaggle enjoys is your safety net for implementation-specific challenges.

WEKA

Waikato Environment for Knowledge Analysis (Weka) is a suite of machine learning tools written in Java. A collection of visualization tools for predictive modelling in a GUI presentation, helping you build your data models and test them, observing the model performances graphically.

TERADATA

A cloud data analytics platform marketing its no code required tools in a comprehensive package offering enterprise-scale solutions. With Vantage Analyst, you don’t need to be a programmer to code complex machine learning algorithms. A simple GUI based system for quick enterprise-wide adoption.

R-Studio

Rstudio is an open-source tool that provides Ide to use R language, and enterprise-ready professional software for data science.

Data mining consists of extracting data from large sets to create more organized and workable data frames that can be manipulated and used to analyze the data set to try to get insight on what the data holds and what it can tell us. There are many algorithms that are used in the realm of data science and analytics. Some important algorithms are:

Classification
Prediction
Clustering
Association
Correlation Analysis

Any Questions??