|  |  |
| --- | --- |
| **Data Analytics** | Analyze raw data to get insights or commonality in the information |
| **Descriptive Analytics** | Describing historical trends in data. Does not make predictions nor informed decisions, merely summarizes data in a meaningful way |
| **Advanced Analytics** | Uses advanced tools to extract data, make predictions, and discover trends. |

Significance

* Optimize and grow businesses through better decision making and products.
* Based on patterns found by analyzing data
* Converts raw data into meaningful information to draw conclusions

Application

* Social media stats
* Cellphone bills show pattern of usage
* Weather monitor - report data instantly
* Player statistics for athletes

**DESCRIPTIVE ANALYTICS**

* Set of techniques that describe what transcribed
* Data queries, reports, data visualization, data-mioning techniques
* Future-prof spreadsheet models

**Data Dashboard**

* Monitor specific aspects of performance related to decision-making
* Collection of information that is updated as new data is available

**Data Mining**

* Use of analytical techniques to understand patterns and relationships in big data

**PREDICTIVE ANALYTICS**

* Techniques that use models from past data to predict future impact of a certain variable
* Linear Regression, Time Series analysis, Data mining, Simulation
* Provides a forecast but not a decision
* Prediction + rule = prescriptive model
* Rule-based models

**BIG DATA**

* Volume (amount)
* Variety (type)
* Velocity (frequency)
* Veracity (quality)
* Value

**TOOLS AVAILABLE FOR DATA ANALYTICS**

* NoSQL
* MapReduce
* Storage
* Servers
* Processing

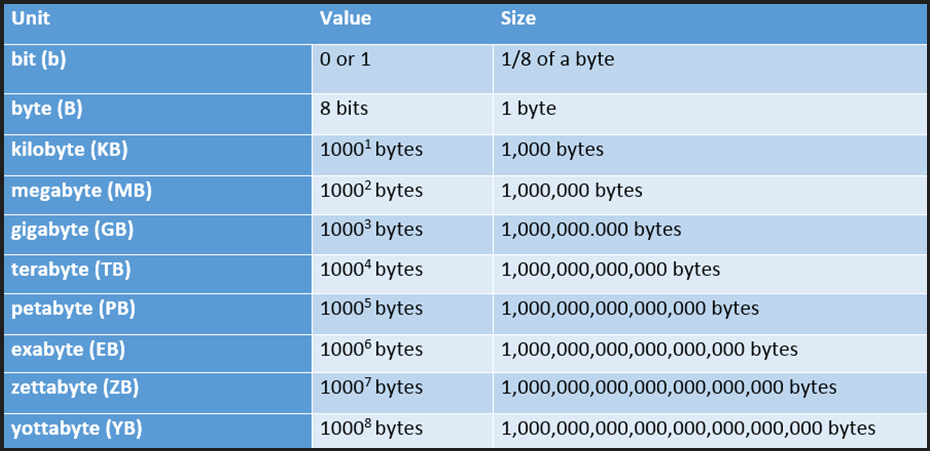
|  |  |
| --- | --- |
| **Descriptive Analysis** | Considers performance indicators and historical data and compares them to a benchmark |
| **Regression Analysis** | Relationship between one or more independent variables and dependent ones |
| **Classification Analysis** | Most applied technique. Combined with logistic regression |
| **Time Series Analysis** | Used for in-depth observation. |



|  |  |
| --- | --- |
| Artificial Neural Networks | It is a system that can improve its structure depending on the information that flows within the network. It considers being highly dependable in forecasting applications and business classification. |
| Decision Trees | A tree-shaped model that describes a regression or classification model. It distributes the data sets into smaller subsets and develops in a similar decision tree simultaneously. |
| Evolutionary Programming | It is domain-independent data analytics techniques that use to examine extensive search space efficiently. |
| Fuzzy Logic | Another data analysis technique depends on the probability that helps in handling the difficulties in data mining techniques. |

**Types of Data**

|  |  |
| --- | --- |
| Qualitative | Quantitative |
| Measurable | Descriptive |
| Collected through measuring things that have a fixed reality | Collected through observation, interviews, recordings |
| Close ended | Open ended |



|  |  |  |
| --- | --- | --- |
| **Parameters** | **Small Data** | **Big Data** |
| *Definition* | Small enough to understand | Complex, difficult to process |
| *Technology* | Traditional | Modern |
| *Database* | SQL | NoSQL |
| *Data Source* | Transactions, Enterprises | Social Media, GPS data, Customer Service |
| *Data Condition* | Ready for analysis, merging not necessary | Unstructured, merging necessary |
| *Data Size* | Spreadsheet is enough | Basta daghan kaayo |
| *Data Purpose* | Data collection | No intended purpose |
| *Velocity* | Aggregation is slow | Aggregation is fast |
| *Scalability* | Vertically | Horizontally |
| *Processing* | Batch-oriented | Batch and stream |
| *Security* | User privileges, hashing, data encryption | Data encryption, access control protocols, cluster network isolation |
| *Location* | Localized, Database | Cloud, offshore, Servers |

