Rosary College of Commerce & Arts
Navelim, Salcete-Goa.

Department of Computer Applications

BITS N BYTES

<u>Principal's Message</u>

The world of data science is evolving every day. Every professional in this field needs to be updated and constantly learning, or risk being left behind.

I am happy that our BCA Department is bringing up this new issue of Bits N Bytes focused on Data Science. Through this electronic platform our students and teachers of Computer Applications can express their views and ideas in this emerging field that could benefit society at large.

I congratulate Asst. Prof Mildred Lemos, Asst. Prof. Sonia Marshal and all who have contributed to Bits N Bytes.

I hope and wish Bits N Bytes will inspire and ignite many minds.

Rev. Dr. Simão R. Diniz
Principal

Introduction to Data Science

-Asst. Prof. Mildred Lemos

Data science is a "concept to unify statistics, **data** analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with **data**. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, computer **science**, and information **science**.

Data Science is primarily used to make decisions and predictions making use of predictive causal analytics, prescriptive analytics (predictive plus decision science) and machine learning.

Predictive causal analytics: A model which can predict the possibilities of a particular event in the future, for example the probability of customers making future credit payments on time.

Prescriptive analytics: A model which has the intelligence of taking its own decisions and the ability to modify it with dynamic parameters. This relatively new field is all about providing advice. The best example for this is Google's self-driving car The data gathered by vehicles can be used to train self-driving cars.

Machine learning for making predictions: A model to determine the future trend for transactional data of a finance company .This falls under the paradigm of supervised learning. It is called supervised because you already have the data based on which you can train your machines. For example, a fraud detection model can be trained using a historical record of fraudulent purchases.

Machine learning for pattern discovery: If you don't have the parameters based on which you can make predictions, then you need to find out the hidden patterns within the dataset to be able to make meaningful predictions. This is nothing but the unsupervised model as you don't have any predefined labels for grouping. The most common algorithm used for pattern discovery is Clustering.

Lifecycle of Data Science
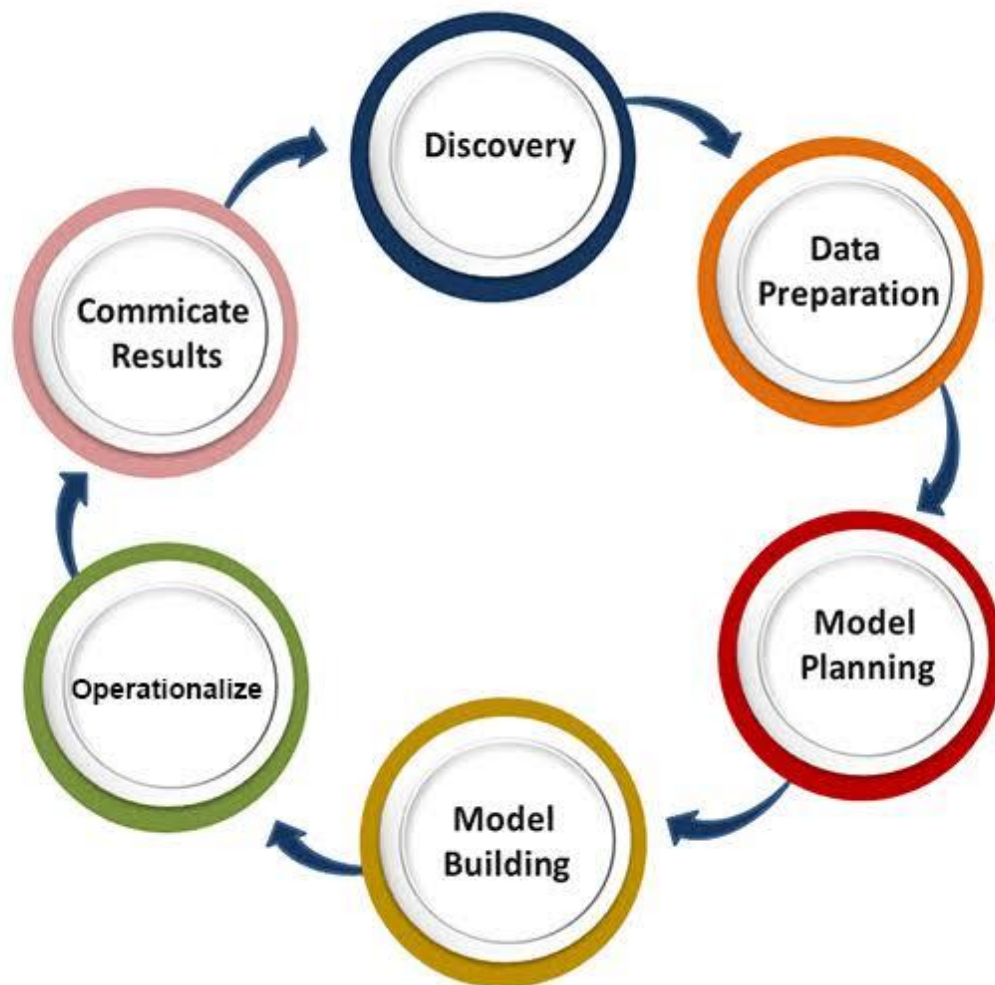
Phase 1—**Discovery:** Before you begin the project, it is important to understand the various specifications, requirements, priorities and required budget.

Phase 2—**Data preparation:** In this phase, You need to explore, preprocess and condition data prior to modeling. Further, you will perform ETLT (extract, transform, load and transform) to get data into the sandbox. R can be used for data cleaning, transformation, and visualization.

Phase 3—**Model planning:** Data Science model planning – In this phase we determine the methods and techniques to draw the relationships between variables. These relationships will set the base for the algorithms.

Phase 4—**Model building:** In this phase, you will develop datasets for training and testing purposes. You will consider whether your existing tools will suffice for running the models or it will need a more robust environment (like fast and parallel processing). You will analyse various learning techniques like classification, association and clustering to build the model.

Phase 5—**Operationalize**: In this phase, you deliver final reports, briefings, code and technical documents.

Phase 6—**Communicate results:** Now it is important to evaluate if you have been able to achieve your goal that you had planned in the first phase. So, in the last phase, you identify all the key findings, communicate to the stakeholders and determine if the results of the project are a success or a failure based on the criteria developed in Phase 1.

<u>Applications of Data Science</u>

- Asst. Prof. Sonia Marshal

Data Science can be used to prevent illegal immigration, identify suspicious activities in crowded areas, predicting locations and movements of nuclear weapons in enemy countries, recognizing and tracking terrorists, detecting violence, flying drones, guiding missiles etc. Using data science and artificial intelligence, upcoming floods in the next 100–500 years can be predicted. These predictions can be used to build dams at correct locations to minimize loss of life and property.

Following are some of the applications of Data Science :-

Healthcare

Procedures such as detecting tumors, artery stenosis, organ delineation employ various different methods and frameworks like MapReduce to find optimal parameters for tasks like lung texture classification. It applies machine learning methods, support vector machines (SVM), content-based medical image indexing, and wavelet analysis for solid texture classification.

Internet Search

Search engines (including Google) make use of data science algorithms to deliver the best result for our searched query in a fraction of seconds.

Targeted Advertising

Starting from the display banners on various websites to the digital billboards at the airports – almost all of them are decided by using data science algorithms.

Website Recommendations

Aren't we all used to the suggestions about similar products on Amazon? They not only help you find relevant products from billions of products available with them but also adds a lot to the user experience.

Advanced Image Recognition

You upload your image with friends on Facebook and you start getting suggestions to tag your friends. This automatic tag suggestion feature uses the face recognition algorithm.

Speech Recognition

Using speech-recognition features, even if you aren't in a position to type a message, you can simply speak out the message and it will be converted to text. Examples of speech recognition products are Google Voice, Siri, Cortana etc.

Gaming

Games are now designed using machine learning algorithms which improve/upgrade themselves as the player moves up to a higher level. In motion gaming also, your opponent (computer) analyses your previous moves and

accordingly shapes up its game. EA Sports, Zynga, Sony, Nintendo, Activision-Blizzard have led gaming experience to the next level using data science.

Traffic prediction in Maps

The time prediction by Google Maps uses the present traffic condition to predict time of travel. But instead it can use predictive analysis to predict traffic at locations at certain time and then use it to predict travel time.