

## Artificial Intelligence and Data Analytics

The future of work is expected to evolve rapidly. Artificial intelligence is going to drive this change. We are already feeling the impact of AI in several products and services that we use almost daily. For example, Google uses an AI tool to selectively show us advertisements. The income tax department is using A.I. with business intelligence to identify people who have not been paying taxes. The world of medicine is undergoing changes with A.I. being used to identify cancerous cells.

Companies are looking to hire data analysts and scientist who can create novel products and solve complex problems. The acute shortage of skilled manpower is being felt globally. Therefore, providing undergraduate students with skills in artificial intelligence and business insights is the need of the hour.

This course is designed to teach students the core concepts of data analytics. It is oriented towards undergraduates who may not necessarily have coding skills. The audience is expected to have the willingness to learn and explore. The focus will be to help students create a portfolio of projects where they can demonstrate the application of data analytics.

This course has 5 modules

**Module 1: Python for Data Analytics** – In this module you will learn how to code in Python. Google, Facebook, Twitter and many more companies use Python to code complex algorithms. They use Python to create models for consumer behaviour, to identify patterns in pictures and to translate one language to another. The coding skills can be learnt in real-time. Like learning any skill, practice is needed to become proficient.

**Module 2: Data Visualization using Power BI** – In the industry, today, there is a lot of data. The challenge is to extract insights that can help propel the business. Power BI is a tool developed by Microsoft that allows users to read, manipulate and extract insights from data.

**Module 3: Statistics for Decision Making** – In this module the basic statistical tools will be taught. These tools help us make decisions using a subset of data also known as sample. This course is about learning the statistical concepts that provide the foundation for A.I. algorithms. The course focusses on the practical aspects of statistics.

**Module 4: Unsupervised Machine Learning** – Pattern recognition has been a topic of significant research. Identifying patterns without looking for any specific pattern is challenging because we do not know what the data structure will reveal. This course focuses on how to implement algorithms that help to identify data structures and patterns.

**Module 5: Supervised Machine Learning** – Business decisions are often made based on predictions about the future. These predictions must be as accurate as possible. In this course the focus is on teaching the A.I. models that predict outcomes based on historical data.

### Evaluation Criteria

A mix of quizzes, homework assignments, projects and exams will be used to assess the learning outcome for this course. Each module will be assessed separately.

### **Module 1: Python for Data Analytics (6 hours)**

- Object Oriented Programming
- Conditionals
- User Defined Functions
- Select In-built Libraries: Numpy, Pandas and Matplotlib

#### Learning Outcome

- Understand how Python is structured and why it is preferred for data analytics
- Write short programmes using Python
- Pre-process data from an excel/csv file before the data can be used in the models

#### Evaluation

- Homework assignment 50%
- Exam 50%

### **Module 2: Data Visualization using Power BI (6 hours)**

- Different visualization tools
- Power Query
- DAX – Data Analytics Expression

#### Learning Outcome

- Understand why Power BI is gaining popularity and is being used by many companies
- Gain an understanding of how data can be processed within Power BI
- Create a story board and extract insights from data

#### Evaluation

- Assignment 20%
- Project 80%

### **Statistics for Decision Making (8 hours)**

- Central Tendency
- Probability Distribution
- Hypothesis Testing
- Analysis of Variance

#### Learning Outcome

- Learn to compute statistical measures
- Use the statistical tools to solve business problems
- Understand the limitations of statistics

#### Evaluation\*

- Assignment 20%
- Project 20%
- Exam 60%

### **Supervised Learning (13 hours)**

- Simple and Multiple Linear Regression
- Classification and Regression Trees
- Logistic Regression

#### Learning Outcome

- Learn to apply models to predict price, churn rate, attrition etc.
- Understand the limitations and difference between models
- Establish an end-to-end framework for data analytics

#### Evaluation

- Assignment 20%
- Project 60%
- Exam 20%

### **Unsupervised Learning (6 hours)**

- Principal Component Analysis
- Clustering Techniques

#### Learning Outcome

- Create models that can extract patterns from data
- Learn to apply dimension reduction techniques to improve predictive capability of the models
- Apply clustering techniques to cluster data based on the attributes in the dataset

#### Evaluation\*

- Assignment 20%
- Project 20%
- Exam 60%