DATA SCIENCE LAB

Data Science being the hottest job skill requirement. This **Lab** focuses on transforming data into valuable insight and applying machine learning and data mining to real-world problems in society and industry.

Computational Resources:

System : Hewllet Packard (HP) Pro Desk 600 SFF

System Count: 69

System Configration:

- HP LED 24" Monitor
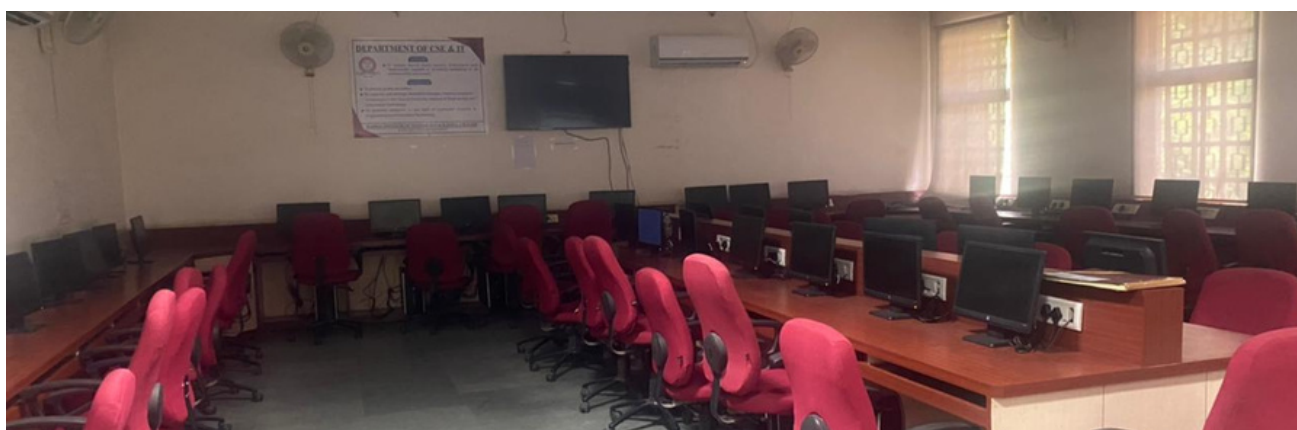- Intel Core i7 9700CPU 3.00 GHZ 8 core
- RAM 16 GB
- HDD 1 TB

Softwares and Utility available:

- **Anaconda3-2018**
- **Python-3.10.4**
- **Mysql5.5**
- **Turbo C**
- **MATLAB r2013b**
- **Microsoft Visual C++**
- **Code Block**
- **SAS**
- **JDK 6u17**
- **R-Studio**

Department Of Computer Science and Engineering

DATA SCIENCE LAB

LAB ETHICS:

- Be on time for your assigned lab session.
- Make Proper entry in log register before taking your assigned system.
- Perform practical's as explained by the facilitator individually, ask for guidance from the facilitator when stuck.
- Aim to finish, at least one practical in a session
- After completion, shut down your systems properly.
- Collect your extensible if any like flash drive, mobiles, notebook, pen extra before you leave your station.
- Arrange the stools and chairs properly before you leave your assigned station and the lab.

In Charge:
Dr Ranjeet Kumar Singh

Physical In Charge:
Mrs Shiva Patel

Department Of Computer Science and Engineering

LIST OF EXPERIMENT:

- Perform Creation, indexing, slicing, concatenation and repetition operations on Python built-in data types: Strings, List, Tuples, Dictionary, Set
- Solve problems using decision and looping statements.
- Apply Python built-in data types: Strings, List, Tuples, Dictionary, Set and their methods to solve any given problem.
- Handle numerical operations using math and random number functions.
- Manipulation of NumPy arrays- Indexing, Slicing, Reshaping, Joining and Splitting.
- Computation on NumPy arrays using Universal Functions and Mathematical methods.
- Import a CSV file and perform various Statistical and Comparison operations on rows/columns.
- Create Pandas Series and DataFrame from various inputs.

LIST OF EXPERIMENT:

- Import any CSV file to Pandas DataFrame and perform the following:
  - Visualize the first and last 10 records
  - Get the shape, index and column details
  - Select/Delete the records(rows)/columns based on conditions.
  - Perform ranking and sorting operations.
  - Do required statistical operations on the given columns.
  - Find the count and uniqueness of the given categorical values.
  - Rename single/multiple columns.

- Import any CSV file to Pandas DataFrame and perform the following:
  - Handle missing data by detecting and dropping/ filling missing values.
  - Transform data using different methods.
  - Detect and filter outliers.
  - Perform Vectorized String operations on Pandas Series.
  - Visualize data using Line Plots, Bar Plots, Histograms, Density Plots and Scatter Plots.

- Use the scikit-learn package in python to implement the regression model and its related methods..

SKILL BASED MINI PROJECT:

Implement the below mentioned models using python programming and related libraries:

● Health Insurance Cost Prediction model.
● Salary Prediction model.
● Loan Amount Prediction model.
● Crop Yield Prediction Model.
● Stock Prediction Model.

COURSE OUTCOME:

After completing the course, the student will be able to:
CO1: Define basic concepts of Data Sciences.
CO2: Illustrate various concepts of python that are used in data sciences.
CO3: Identify various methods for the representation and manipulation of\
       vectors.
CO4: Analysis the data for applying various statistical modelling
       approaches.
CO5: Identify hidden patterns in data and transform it using data science
       techniques.
CO6: Apply regression techniques to solve real world problems.

LIST OF EXPERIMENT- MACHINE LEARNING

- Perform Creation, indexing, slicing, concatenation and repetition operations on Python built- in data types: Strings, List, Tuples, Dictionary, Set
- Solve problems using decision and looping statements.
- Apply Python built-in data types: Strings, List, Tuples, Dictionary, Set and their methods to solve any given problem.
- Manipulation of NumPy arrays- Indexing, Slicing, Reshaping, Joining and Splitting.
- Computation on NumPy arrays using Universal Functions and Mathematical methods.
- Import a CSV file and perform various Statistical and Comparison operations on rows/columns.
- Create Pandas Series and DataFrame from various inputs.
- Import any CSV file to Pandas DataFrame and perform the following:
  - Visualize the first and last 10 records
  - Get the shape, index and column details
  - Select/Delete the records(rows)/columns based on Conditions
  - Perform ranking and sorting operations.

LIST OF EXPERIMENT- MACHINE LEARNING

- Do required statistical operations on the given columns.
- Find the count and uniqueness of the given categorical values.
- Rename single/multiple columns.

- Import any CSV file to Pandas DataFrame and perform the following:
  - Handle missing data by detecting and dropping/ filling missing values.
  - Transform data using different methods.
  - Detect and filter outliers.
  - Perform Vectorized String operations on Series.

- Use scikit-learn package in python to implement following machine learning models to solve real world problems using open source datasets:
  - Linear Regression model.
  - Multi-linear regression model.
  - Decision tree classification model.
  - Random forest model.
  - SVM model.
  - K-means clustering model.

COURSE OUTCOME:

After completing the course, the student will be able to:

CO1: Understand basic concepts of Machine Learning.

CO2: Illustrate various data cleansing and pre-processing techniques.

CO3: Separate regression and classification problems.

CO4: Analysis the data for applying various supervised and unsupervised machine learning techniques.

CO5: Evaluate different machine learning models based on performance parameters.

CO6: Design machine learning solutions for real world problems.