DATA SCIENCE: CAREER OF THE FUTURE

INTRODUCTION TO DATA SCIENCE

SANJAY RAJVANSHI

SCHEDULE

Session	Date	Time	Topic
1	Sep 25	7:00 pm – 8:00 pm	Introduction to data science and associated tools.
2	Oct 2	7:00 pm – 8:00 pm	Introduction to Python. Learn how to use Python for data analysis. Python is simple, yet powerful language that is often used in data science.
3	Oct 9	7:00 pm – 8:00 pm	Data wrangling with Python. Learn how to gather data and make it useful for analysis.
4	Oct 16	7:00 pm – 8:00 pm	Data visualization and analysis with Python. Learn how to create useful visualizations to aid in the analysis of the data.
5	Oct 23	7:00 pm – 8:00 pm	Brief introduction to artificial intelligence and machine learning. Get a peek into how to make data based predictions.

Note: All classes are on Wednesdays.

SESSION 4 – RECAP

- Data visualization and analysis with Python.
- Created useful visualizations to aid in the analysis of the data.
- Created and customized various types of graphs.
- Learnt some statistical techniques.

SESSION 5: INTRODUCTION TO MACHINE LEARNING

SESSION 5 – AGENDA

- More data visualization and analysis with Python.
- Learn more statistical techniques as they come up.
- Brief introduction to artificial intelligence and machine learning.
- Get a peek into how to make data based predictions.

DATA SCIENCE SOLUTION LIFECYCLE

- Data Science solution lifecycle (iterative):
  - Problem identification
  - Identify data
  - Clean, transform data
  - **Analyze, visualize**
  - **Identify algorithm(s)**
  - Implement
  - Maintain and support

ARTIFICIAL INTELLIGENCE, MACHINE LEARNING

- Artificial Intelligence [1]
  - "Artificial intelligence is the science and engineering of making computers behave in ways that, until recently, we thought required human intelligence." – Andrew Moore, Former-Dean of the School of Computer Science at Carnegie Mellon University.
- Machine Learning [1]
  - "Machine learning is the study of computer algorithms that allow computer programs to automatically improve through experience." – Tom M. Mitchell, Former Chair of the Machine Learning Department at Carnegie Mellon University.
  - Two broad classifications:
    - Supervised Learning
    - Unsupervised Learning

MACHINE LEARNING

- Supervised Learning [2]
  - Given a dataset, the correct result or the "ground truth" is already known i.e. there is a relationship between the input and the output.
  - Types of approaches
    - Regression – predict results within a continuous output
    - Classification – predict results in a discrete output
- Unsupervised Learning [2]
  - Given a dataset, the correct result is NOT known.
  - Derive structure from data based on relationships among the variables in the data.
  - Types of approaches:
    - Clustering – automatically group data into groups related by different variables.

EXERCISE

- Create a Python file with name "S5-Exx".
- We will cover some of the topics in previous slides in this exercise working directly in the Jupyter notebook.

SESSION 5 – HOME WORK

- Research on the internet and identify Machine Learning problems – at least for each of the approaches i.e. for Regressions, Classification, Clustering.
- Research Neural Network approach.

CLASS – RECAP

- Introduction to data science and associated tools.
- Introduction to Python.
- Data wrangling with Python.
- Data visualization and analysis with Python.
- Brief introduction to artificial intelligence and machine learning. A peek into how to make data based predictions.

REFERENCES

Note: No sign-up required unless you want to be part of a meeting group or take a course.

1. *Machine Learning in Python* – https://scikit-learn.org/stable/#
2. *Machine Learning in Python (more details)* – https://scikit-learn.org/stable/user_guide.html
3. *Machine Learning in Python (examples)* – https://scikit-learn.org/stable/auto_examples/index.html
4. *Meeting groups* – Meetup.com
5. *Machine Learning vs. AI, What are the Important Differences Between the Two?* - https://medium.com/datadriveninvestor/differences-between-ai-and-machine-learning-and-why-it-matters-1255b182fc6
6. *Machine Learning, Coursera course by Stanford University, Prof. Andrew Ng* – https://www.coursera.org/learn/machine-learning