CS677
Data Science with Python

Suresh Kalathur, PhD

kalathur@bu.edu

Overview

- *Topics*
  - *http://kalathur.com/courses/?course_id=cs677_19_fall*
- Class participation & Discussions (20%)
- Programming Assignments (20%)
- Final Exam (30%)
- Term Project (30%)

Topics

- Review of Python
- Numpy and Pandas
- Data Visualization
- Data Wrangling, Aggregation
- Regression, Time Series
- Predictive Analytics, Text Analytics
- Applications

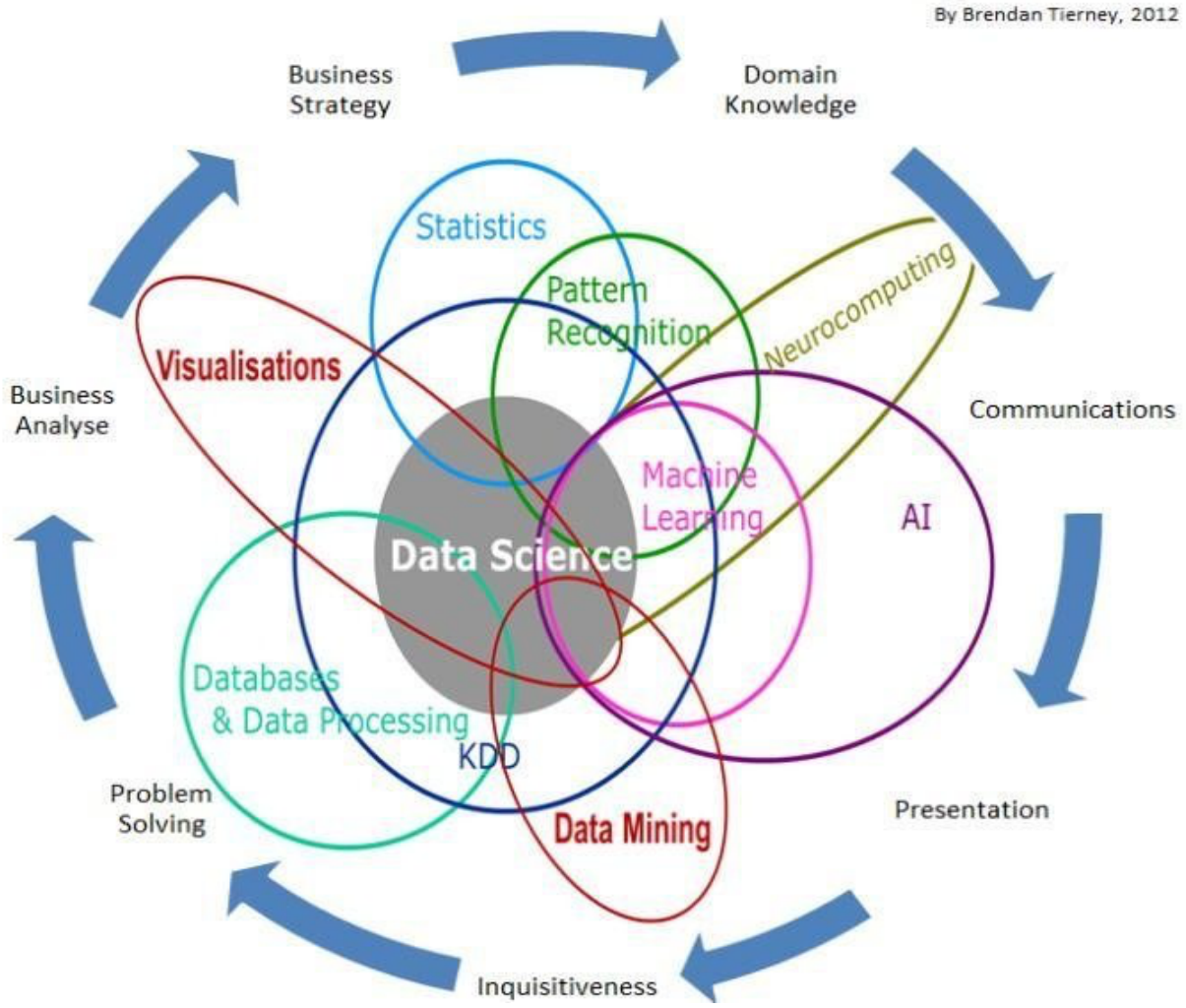
Software Setup

- Python 3.x
  - https://www.python.org/downloads/
- Jupyter
  - python3 -m pip install --upgrade pip
  - python3 -m pip install jupyter
- For MACs
  - /Applications/Python 3.7/Install\ Certificates.command
- Run Jupyter
  - jupyter notebook

Initial Setup

- Install Python Modules
  - python3 –m pip install <x>
    - numpy pandas
    - scipy scikit-learn statsmodels
    - matplotlib seaborn plotly
    - Pillow

Data Science Is Multidisciplinary



Data Analyst

- Descriptive statistics, visualize data, analyze conclusions
- Knowledge of mathematical statistics
- Fluent with R and Python
- Data wrangling tasks
- Understand big data technologies (Hadoop, Spark, etc.)

Machine Learning

- Using algorithms on data, learn from it, and forecast future trend
- Traditional machine learning
  - Statistical and predictive analysis
  - Linear regression, logistic regression (binary classification), linear discriminant analysis, decision trees, Naïve Bayes, k-Nearest Neighbors, Support Vector Machines, etc.

Data Science

- Amalgamation from multiple disciplines
  - Data analytics
  - Machine learning
  - Predictive analytics
  - Business analytics
  - Software engineering
  - Data engineering, etc.

Lecture

- Software Setup
- Review of Python