Course 1: Intelligent Data Visualization – Getting to Know your Data

Course Overview:

In this course, students will be introduced to the various types of data commonly utilized in Artificial Intelligence and Machine Learning problems. In addition, the students will apply standard processes commonly found in data extraction, data transformation and data loading methods commonly utilized in AI applications. Students will construct a data ingestion pipeline using a data warehouse and prep data for use in a predefined Machine Learning problem.

Course Outline

This course will cover the following topics:

* Data Extraction
  + Understand full scope of ETL (Extraction, Transform, Load) for data ingestion pipeline
  + Data migration
  + Data Combination
  + Structured Data
    - Full Extraction
    - Incremental Extraction
  + Unstructured Data focusing on data preparation
  + How to address common issues in the extraction process
  + Highlight data extraction tools:
    - Batch processing tools
    - Open Source Tools
    - Cloud-Based Tools
* Data Transformation
  + Identify Variable Types
  + Feature Transformation
  + Power Transform
  + Difference Transform
  + Standardization
  + Normalization
  + Binning
* Data Loading
  + Challenges of Data loading
  + Data Loading:
    - Batch processing tools
    - Open Source Tools
    - Cloud-Based Tools
* Google BigQuery, Snowflake, and Amazon Redshift or S3???

Learning Outcomes

The Intelligence Factory (TIF) Intelligent Data Visualization course is comprehensive foundation course for the fields of data science, artificial intelligence, and machine learning. The ETL knowledge and skills learning in this course are transferrable to all other TIF courses and certifications as well as prepare the student with industry-relevant skills and tools to understand and manipulate various types of data.

Successful completion of this course entitles the student to The Intelligence Factory Skillset Certification in the following areas:

* Data Visualization (Tableau, t-SNE)
* Feature Engineering

This course also provides you the skills and concepts needed to complete the TIF Certification One for a Machine Learning Engineer.

Duration

30 hours/ 5 weeks/ 6 hours per week/3 hours for two nights

3 hours per night

3 lessons per night (

Week 1

hour 1:

A: Lesson I

15 min explanation and ‘show’

15 min ‘do’

B: Lesson II

15 min explanation and ‘show’

15 min ‘do’

hour 2

A: Lesson III

15 min explanation and ‘show’

15 min ‘do’

B: Lesson IV

15 min explanation and ‘show’

15 min ‘do’

hour 3

A:

15 min explanation and ‘show’

15 min ‘do’

B:

15 min explanation and ‘show’

15 min ‘do’

Resources and Materials

* Access to Learning Lab software
* GitHub acct
* Slides and resources will be provided

Deliverables

Pre-work: Python Primer (Pycharm and Jupyterhub)

Week 1:

Session 1

Session 2

Lecture/Walkthrough (2 hours):

ML/AI – types of ML, what is AI?

Introduction to ETL

Datatypes

SQL

Class Scenario (2 hours): API exercise or web scraping

Week 2: Extraction (think about Kafka or API, web scraping)

<https://towardsdatascience.com/creating-a-dataset-using-an-api-with-python-dcc1607616d>

pull data from an api and insert it into MySQL

* Data Extraction

connect to msql and write some quires to extract the data – similar to xtol – use weather or energy data for this

* + Understand full scope of ETL (Extraction, Transform, Load) for data ingestion pipeline
  + Data migration
  + Data Combination
  + Structured Data
    - Full Extraction
    - Incremental Extraction
  + Unstructured Data focusing on data preparation
  + How to address common issues in the extraction process
  + Highlight data extraction tools:
    - Batch processing tools
    - Open Source Tools
    - Cloud-Based Tools

Week 3: Transformation

Pandas

* Data Transformation
  + Identify Variable Types
  + Feature Transformation
  + Power Transform
  + Difference Transform
  + Standardization
  + Normalization
  + Binning

Week 4: Loading

Pandas

* Data Loading
  + Challenges of Data loading
  + Data Loading:
    - Batch processing tools
    - Open Source Tools
    - Cloud-Based Tools

Week 5: Capstone