## 1. Course Code

2262

## 2. Course Title

G52e: Data Science Exercises

## 3. Teacher

HAMIDULLAH, Sokout

## 4. Term

Spring 2

## 5. Course Requirements (Courses / Knowledge for this course) and Important Information

Data Science Course (2261)

## 6. Course Overview and Objectives

The major goal of Data Science is to solve companies (institutions, government, private sectors, .. etc) problems using available data. Data Science entails everything that has something to do with data such as: Collecting the data, Cleaning up the data, Visualizing data, Analyzing it, and Creating predictive models based on the data. Playing with data will pave the way to gain insights that could help us with good decision making.

In this course we will start from the very basics (pre-processing data), we will build up your skills and soon you will be able to solve advanced statistics tasks using machine learning algorithms in Python such as CLUSTERING, REGRESSION, CLASSIFICATION.

The participants will also learn the concept of Nerual Networks (Deep Learning) and how to apply it in real business problem. For this year, we added the fundamentals concept of Big Data and the Hadoop ecosystem.

## 7. Course Outline

1 Course Orientation and Data Preprocessing with Python
2 Data Preprocessing Exercises with Machine Learning (Regression)
3 Machine Learning Algorithmes (Classification)
4 Classification Exercises
5 Model Selection
6 Exercises
7 Machine Learning Algorithms (Clustering)
8 Clustering Exercises
9 Integration Concept (Case Study)
10 Integration Exercises (In-Class Case Study)
11 Deep Learning Concepts
12 Exercises with Deep Learning Algorithem
13 Big Data (Hadoop Ecosystem) -1
14 Big Data (Hadoop Ecosystem) -2
15 Students Presentation
16 Students Presentation

## 8. Textbooks (Required Books for this course)

## 9. Reference Books (optional books for further study)

(1) Machine Learning with Python Cookbook, ISBN: 9781491989388.
(2) Advanced Data Analytics using Python, ISBN: 978-1-4842-3449-5.
(3) Available resources on Internet

## 10. Course Goals (Attainment Targets)

(1) Able to preprocess and engineering data using Python
(2) Gain in-depth familiarity with various Machine Learning algorithms (supervised learning algorithms and unsupervised learning algorithms), as well as Deep Learning
(3) Able to Implement machine learning and deep learning algorithms to real-world problems, and rigorously evaluate their performance using different methods.
(4) Understand the concpet of Integration Programming and implement in real-world problems.
(5) Understand the core components of the Hadoop ecosystem and will explore the architecture, components and functionalities of Hadoop.
(6)
(7)
(8)

## 11. Correspondence relationship between Educational goals and Course goals

| Educational goals of the school | | | Course Goals |
|---|---|---|---|
| High level ICT skills | Basic academic skills | | |
| | Specialized knowledge and literacy | | (1) (2) (3) (4) (5) |
| Human skill (Tankyu skill) | Ability to continually improve own strengths | | (1) (2) (3) (4) (5) |
| | Ability to discover and resolve the problem in society | Problem setting | (1) (2) (3) (4) (5) |
| | | Hypothesis planning | (1) (2) (3) (4) (5) |
| | | Hypothesis testing | (1) (2) (3) (4) (5) |
| | | Practice | (1) (2) (3) (4) |
| | Fundamental Competencies for Working Persons | Ability to step forward | (3) (4) (5) |
| | | Ability to think through | (3) (4) (5) |
| | | Ability to work in a team | (3) (4) (5) |
| Professional ethics | | | |

## 12. Evaluation

| Goals | Evaluation method & point allocation | | | | | |
|---|---|---|---|---|---|---|
| | Examination | Quiz | Reports | Presentation | Deliverables | Other |
| (1) | ◯ | ◯ | | ◯ | ◯ | |
| (2) | ◯ | ◯ | | ◯ | ◯ | |
| (3) | ◯ | ◯ | | ◯ | ◯ | |
| (4) | ◯ | ◯ | | ◯ | ◯ | |
| (5) | | | | ◯ | ◯ | |
| (6) | | | | | | |
| (7) | | | | | | |
| (8) | | | | | | |
| Allocation | 20 | 20 | | 20 | 40 | |

## 13. Evaluation Criteria

| | |
|---|---|
| Examination | A case study will be assighned to participants. They are tasked to take advantage of the public dataset and build ML model for real research problem. In this case study, the participants will apply the steps of data preprocessing, developing a ML model and applyting approperiate avaluation metrics and finally deploy the most accurate model to a responsive web application along with approperiate visualtization. |
| Quiz | Weekly short in-class quizzes will be administered to reinforce understanding of the week's topics. These quizzes will consist of coding-based questions and are designed to encourage active engagement, peer discussion, and timely feedback. They support continuous learning and help identify areas that may need further clarification. |
| Reports | |

| Presentation | In the final presentation, students will be asked to work on a real-world problem using Machine Learning algorithm. Participants are tasked to take advantage of the public dataset and build AI-based solution.<br><br>The evaluation will be based on participation in the group presentation, the structure of presentation, the relavance of the argument, time managment and relation to the course contents. |
|---|---|
| Deliverables | Individual and group assignment will be assigned for the students with focus on learning goals (1,2,3, and 4). The evaluation will be based on how the students understand the exercises and participation. |
| Other | |

## 14. Active Learning

| Hourly percentage of active learning within the whole class time | 80% |
|---|---|

| | | |
|---|---|---|
| 1 | Active learning such as problem solving assignment using the knowledge and skills acquired in class. | All the time |
| 2 | Active learning such as group works and discussions. | All the time |
| 3 | Outcome presentations and feedbacks. | Sometimes |
| 4 | Students actively make decisions on how the class should be conducted. | Not at all |

## 15. Notes

This course is like a jouney by attending it you will benefit and might enjoy. Please bring your own computer in the class.

## 16. Course plan

(Notice) This plan is tentative and might be changed at the time of delivery

| **Lessen 1: Data Preprocessing + Regression with Python** | **Lecture + Exercises** |
|---|---|

1. Apply dataset preparation mechanisms using Sciketlearn
2. Apply Machine Learning techniques (Regression)
3. Apply model evaluation techniques

| **Lessen 2: Data Preprocessing + Regression with Python Exercises** | **Exercises** |
|---|---|

1. Exercises and Practices with Regression Problems

| **Lessen 3: Machine Learning Algorithmes (Classification Concept)** | **Lecture + Exercises** |
|---|---|

1. Apply Machine Learning techniques (Classification)
2. Apply model evaluation techniques

| **Lessen 4: Machine Learning Algorithmes (Classification Exercises)** | **Exercises** |

1. Exercises and Practices with Classification problems

| **Lessen 5: Model parameters (hyperparameters) selection** | **Lecture + Exercises** |

1. Model Selection
    a. K-fold cross validation
    b. Grid Search
    c. XG-Boost
    d. Cat-Boost

| **Lessen 6: Model Selection Exercises** | **Exercises** |

1. Parameter tuning for model optimization

| **Lessen 7: Machine Learning Algorithmes (Unsupervised Concept)** | **Lecture + Exercises** |

1. Apply Machine Learning Algorithmes (Unsupervised ML)
2. K-Means Concepts and Exercises with real dataset

| **Lessen 8:  Machine Learning Algorithmes (Unsupervised Concept)** | **Exercises** |

1. Exercises and Practices Unsupervised ML with read dataset

| **Lessen 9: Integration + Model deployment (Case Study)** | **Exercises** |

1. Integration Concept
    a. Python
    b. My-SQL
    c. Tableau

| **Lessen 10: Integration + Model deployment Exercises (Case Study)** | **Exercises** |

1. Devlopment Model with Integration Concept

| **Lessen 11: Deep Learning Algorithms Concept and Exercises** | **Lecture + Exercises** |

1. Deep Learning Concpet
2. Exercises with Deep Learning Algorithm
    a. ANN
    b. CNN

| **Lessen 12: Deep Learning Algorithms Concept and Exercises** | **Exercises** |
| --- | --- |

    1. Exercises with Deep Learning Algorithm

| **Lessen 13: Hadoop Ecosystem-1** | **Lecture** |
| --- | --- |

    1. Introduction to Big Data & Hadoop
        a. Hadoop Ecosystem: HDFS, MapReduce, YARN
        b. ETL concept and how it works
        c. NoSQL databases

| **Lessen 14: Hadoop Ecosystem-1** | **Lecture + Exercises** |
| --- | --- |

    1. Data Processing with Hadoop & Spark
    2. Data Storage & Management with Hadoop

| **Lessen 15: Student Presentation** | **Presentation(90 min)** |
| --- | --- |

Students Presentation

| **Lessen 16 : Student Presentation** | **Presentation (90 min)** |
| --- | --- |

Students Presentation