| | Universitas Negeri Surabaya<br>Faculty of Mathematics and Natural Sciences<br>Data Science Undergraduate Study Program | Document Code |
|---|---|---|

# SEMESTER LEARNING PLAN

| Courses | CODE | Course Family | Credit Weight | | | SEMESTER | Compilation Date |
|---|---|---|---|---|---|---|---|
| Text Processing | 4920202030 | | T=3 | P=0 | ECTS=4.77 | 3 | July 29, 2023 |

| AUTHORIZATION | SP Developer | Course Cluster Coordinator | Study Program Coordinator |
|---|---|---|---|
| | Riskyana Dewi Intan Puspitasari, M.Kom | Dr. Elly Matul Imah., M.Kom | Yuliani Puji Astuti, S.Si., M.Si. |

**Learning model**: Project Based Learning

**Program Learning Outcomes (PLO)**

**PLO study program that is charged to the course**

| | |
|---|---|
| PLO-9 | Able to apply data science principles to solve problems |
| PLO-16 | Mastering data science theories and concepts |

**Program Objectives (PO)**

| | |
|---|---|
| PO - 1 | Explains the concept of Language Modeling in text processing |
| PO - 2 | Able to represent linguistic knowledge at the level of morphological, syntactic and semantic representation |
| PO - 3 | Able to extract text data from digital sources and process it using pre-processing techniques, feature extraction and text classification |
| PO - 4 | Able to design problem solving on text data using related text data processing |

**PLO-PO Matrix**

| P.O | PLO-9 | PLO-16 |
|---|---|---|
| PO-1 | | |
| PO-2 | | |
| PO-3 | | |
| PO-4 | | |

**PO Matrix at the end of each learning stage (Sub-PO)**

| P.O | Week | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| PO-1 | | | | | | | | | | | | | | | | |
| PO-2 | | | | | | | | | | | | | | | | |
| PO-3 | | | | | | | | | | | | | | | | |
| PO-4 | | | | | | | | | | | | | | | | |

**Short Course Description**: This course is a project-based course that studies basic techniques for processing text data. The course will introduce the concepts of language morphology, text representation, pre-processing, feature extraction to obtain information such as similarity and text clustering. Topics covered include: language morphology, string representation, regex, tokenization, text pre-processing, Bag of Words, TF-IDF, word similarity, word clustering, and web scraping. Students will create group projects to apply text processing theories and concepts to problems in the field of Data Science.

**References**

**Main :**


1. Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019


**Supporters:**


1. Daniel Jurafsky & James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2nd Edition, Prentice Hall, 2008.


**Supporting lecturer**: Dr. Elly Matul Imah, M.Kom.
Riskyana Dewi Intan Puspitasari, M.Kom.

| Week- | Final abilities of each learning stage (Sub-PO) | Evaluation | | Help Learning, Learning methods, Student Assignments, [ Estimated time] | | Learning materials [ References ] | Assessment Weight (%) |
|---|---|---|---|---|---|---|---|
| | | Indicator | Criteria & Form | Offline ( *offline* ) | Online ( *online* ) | | |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 1 | Understand the concept of language morphology starting from phonemes, morphemes, lexemes, syntax, and context | 1. Explain the concept of human language morphology<br>2. Explain the concept of phoneme<br>3. Explain the concepts of morpheme and lexeme<br>4. Explain the concept of syntax<br>5. Explain the concept of context | **Criteria:**<br>Non-Test Assignments<br><br>**Form of Assessment** :<br>Participatory Activities | Collaborative Learning (Lectures, discussions and questions and answers)<br>3 X 50 | | **Material:**<br>Language Concepts, Phoneme, Morpheme and Lexeme, Syntax, Context<br>**Bibliography:**<br>*Daniel Jurafsky & James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2nd Edition, Prentice Hall, 2008 .* | 2% |
| 2 | Explains string representation, string operations, regex, and text data visualization in programming | 1. Explains string representation.<br>2. Explains operations on strings<br>3. Explain indexing and slicing techniques<br>4. Explains methods on strings<br>5. Explains string formatting<br>6. Explains regex on strings | **Criteria:**<br>Non-Test Assignments<br><br>**Form of Assessment** :<br>Participatory Activities | Collaborative Learning (Lectures, discussions and questions and answers)<br>3 X 50 | | **Material:**<br>String representation, String operations, Indenxing and slicing, String method, String formatting, String regex<br>**Library:**<br>*Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 2% |
| 3 | Perform text data pre-processing techniques | 1. Removes HTML tags<br>2. Perform tokenization<br>3. Remove stop words<br>4. Overcoming spelling errors in text | **Criteria:**<br>Non-Test Assignments<br><br>**Form of Assessment** :<br>Participatory Activities | Collaborative Learning (Lectures, discussions and questions and answers)<br>3 X 50 | | **Material:**<br>Removing HTML tags, Tokenizing, Removing stopwords, Overcoming spelling errors in text<br>**References:**<br>*Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 2% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **4** | Perform text data pre-processing techniques | 1.Stemming.<br>2.Performing Lemmatization<br>3.Tagging<br>4.Chunking<br>5.Perform Parsing | **Criteria:**<br>Task<br><br>**Form of Assessment** :<br>Participatory Activities, Practice/Performance | Collaborative Learning (Lectures, discussions and questions and answers)<br>3 X 50 | | **Material:** Stemming, Lemmatization, Tagging, Chunking, Parsing<br>**Library:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 10% |
| **5** | Perform feature extraction techniques on text data | 1.Do the Bag of Words technique<br>2.Carrying out the Bag of N-grams technique<br>3.Perform Parsing | **Criteria:**<br>Task<br><br>**Form of Assessment** :<br>Participatory Activities | Collaborative Learning (Lectures, discussions and questions and answers)<br>3 X 50 | | **Material:** Bag of Words, Bag of N-gram Words<br>**Library:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 2% |
| **6** | Perform feature extraction techniques on text data | 1.Perform the TF-IDF technique<br>2.Carrying out the similarity features technique | **Criteria:**<br>Non-Test<br><br>**Form of Assessment** :<br>Participatory Activities | Collaborative Learning (Lectures, discussions and questions and answers)<br>3 X 50 | | **Material:** TF-IDF, Similarity Features<br>**Library:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 2% |
| **7** | Perform feature extraction techniques on text data | 1.Performing the Word2Vec technique<br>2.Do the Glove technique<br>3.Performing the FastText technique | **Criteria:**<br>Non-Test<br><br>**Form of Assessment** :<br>Participatory Activities | Collaborative Learning (Lectures, discussions and questions and answers)<br>3 X 50 | | **Materials:** Word2Vec, Glove, FastText<br>**Libraries:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 2% |
| **8** | | Able to answer questions accurately and correctly | **Criteria:**<br>Written Test Exam<br><br>**Form of Assessment** :<br>Practice/Performance, Test | Midterm Exam (UTS)<br>100 minutes | | **Material:** All material before UTS<br>**Library:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 10% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **9** | Perform text similarity on text data | 1. Perform text similarity between terms and documents<br>2. Finds the most relevant documents using distance techniques<br>3. Create a system recommendation from text similarity techniques | **Criteria:**<br>Non-Test | Collaborative approach (discussion and expository) Discussion 3x50 | | **Material:** Character vectorization, Hamming Distance, Manhattan Distance, Euclidean Distance, Cosine Distance, Document Similarity, Recomender System<br>**Library:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 2% |
| **10** | Perform text clustering on text data | 1. Perform text clustering using a hierarchical clustering model<br>2. Perform text clustering using the distribution clustering model<br>3. Perform text clustering using the density clustering model | **Criteria:**<br>Non-Test<br><br>**Form of Assessment** : Participatory Activities | Collaborative approach (discussion and expository) Discussion 3x50 | | **Material:** hierarchical clustering model, distribution clustering model, density clustering model<br>**References:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 2% |
| **11** | Performing Web Scraping (HTTP) | 1. Introduction to the Beautifulsoap library<br>2. Get to know the methods and attributes of beautifulsoap<br>3. Crawling website pages<br>4. Save data scraping results | **Criteria:**<br>Non-Test<br><br>**Form of Assessment** : Participatory Activities, Practice/Performance | Collaborative approach (discussion and expository) Discussion 3x50 | | **Material:** Beautiful questions, Website Layout, Crawling<br>**Library:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 2% |
| **12** | Performing Web Scraping (API) | 1. Get to know the API concept<br>2. Perform parsing using JSON<br>3. Crawling website pages via API<br>4. Save data scraping results | **Criteria:**<br>Test<br><br>**Form of Assessment** : Participatory Activities, Practice/Performance | Collaborative approach (discussion and expository) Discussion 3x50 | | **Material:** Scrapping using HTTP, API, JSON<br>**References:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 10% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **13** | Implement data processing techniques on real problems | 1.Determine the topic of the problem<br>2.Arranging a team and project work schedule<br>3.Create project problem solving designs | **Criteria:**<br>Presentation<br><br>**Forms of Assessment** :<br>Participatory Activities, Project Results Assessment / Product Assessment, Practices / Performance | Participatory Activities, Practice / Performance 3x50 | | **Material:** All material<br>**Library:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 10% |
| **14** | Implement data processing techniques on real problems | 1.Implement project results<br>2.Realizing approved project results | **Criteria:**<br>Presentation<br><br>**Form of Assessment** :<br>Participatory Activities, Project Results Assessment / Product Assessment | Participatory Activities, Practice / Performance 3x50 | | **Material:** All material<br>**Library:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 5% |
| **15** | Implement data processing techniques on real problems | Presentation of progress and report on group work achievements | **Criteria:**<br>Presentation<br><br>**Form of Assessment** :<br>Participatory Activities, Project Results Assessment / Product Assessment | Participatory Activities, Practice / Performance 3x50 | | **Material:** All material<br>**Library:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 9% |
| **16** | Final Semester Examination (UAS) | Presentation and question and answer | **Criteria:**<br>Participatory Activities, Project Results Assessment / Product Assessment, Portfolio Assessment<br><br>**Forms of Assessment** :<br>Participatory Activities, Project Results Assessment / Product Assessment, Portfolio Assessment | Presentation and questions and answers 3x50 | | **Material:** All material<br>**Library:** *Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019* | 30% |

**Evaluation Percentage Recap: Project Based Learning**

| No | Evaluation | Percentage |
|---|---|---|
| 1. | Participatory Activities | 45.33% |
| 2. | Project Results Assessment / Product Assessment | 20.33% |
| 3. | Portfolio Assessment | 10% |
| 4. | Practice / Performance | 19.33% |
| 5. | Test | 5% |
| | | 99.99% |

**Notes**

1. **Learning Outcomes of Study Program Graduates (PLO - Study Program)** are the abilities possessed by each Study Program graduate which are the internalization of attitudes, mastery of knowledge and skills according to the level of their study program obtained through the learning process.
2. **The PLO imposed on courses** are several learning outcomes of study program graduates (CPL-Study Program) which are used for the formation/development of a course consisting of aspects of attitude, general skills, special skills and knowledge.
3. **Program Objectives (PO)** are abilities that are specifically described from the PLO assigned to a course, and are specific to the study material or learning materials for that course.
4. **Subject Sub-PO (Sub-PO)** is a capability that is specifically described from the PO that can be measured or observed and is the final ability that is planned at each learning stage, and is specific to the learning material of the course.
5. **Indicators for assessing** ability in the process and student learning outcomes are specific and measurable statements that identify the ability or performance of student learning outcomes accompanied by evidence.
6. **Assessment Criteria** are benchmarks used as a measure or measure of learning achievement in assessments based on predetermined indicators. Assessment criteria are guidelines for assessors so that assessments are consistent and unbiased.

Criteria can be quantitative or qualitative.

7. **Forms of assessment:** test and non-test.
8. **Forms of learning:** Lecture, Response, Tutorial, Seminar or equivalent, Practicum, Studio Practice, Workshop Practice, Field Practice, Research, Community Service and/or other equivalent forms of learning.
9. **Learning Methods:** Small Group Discussion, Role-Play & Simulation, Discovery Learning, Self-Directed Learning, Cooperative Learning, Collaborative Learning, Contextual Learning, Project Based Learning, and other equivalent methods.
10. **Learning materials** are details or descriptions of study materials which can be presented in the form of several main points and sub-topics.
11. **The assessment weight** is the percentage of assessment of each sub-PO achievement whose size is proportional to the level of difficulty of achieving that sub-PO, and the total is 100%.
12. TM=Face to face, PT=Structured assignments, BM=Independent study.