IMPLEMENTATION OF DATA ANALYTICS IN PYTHON PROGRAMMING LEARNING ASSISTANCE SYSTEM

By

Noora Aulia Hidayat

NIM. 2041720046

Supervisor:

1. **Yan Watequlis Syaifudin, ST., M.MT., Ph.D.**
2. **Retno Damayanti, S.Pd., M.T.**

INTRODUCTION

The demand for Data Analytics in the business world is increasing. In 2023, there were 1,072 job postings related to the most sought-after skills, techniques, and degrees, as well as the current job market conditions on LinkedIn (Yosifova, 2023).

In programming learning without validation, students find it difficult to measure how well they understand the material. Similarly, identifying errors and finding mistakes.

To facilitate this, The lecturer have taken the initiative to develop a web-based learning platform equipped with automated code validation (Watequlis Syaifudin, 2021).

Research problem

1. How to measure the level of understanding of students in the Python Assistance System for Data Analytics Learning website?
2. How can this website assist students in self-learning Python programming for Data Analytics?

Objectives of the Research

1. Implementing Data Analytics Learning in Python Programming Assistance System
2. Assisting students in independently learning Python programming for Data Analytics.

Scoup of Research

1. Learning platform based on website
2. The platform using Google Colaboratory
3. Users are students of Politeknik Negeri Malang
4. Utilizing the Python programming languages
5. Learning materials related to Data Analytics include load data, data cleaning, data manipulation, data data visualization.

Previous Research

- the Implementation of Basic Programming Learning in Python. This research is based on a website for easy access. The system is designed to facilitate students in independently learning the basics of Python programming. User-inputted code will be evaluated using Codewars (Yulvarisma, 2022),
- Python Programming Learning Assistance System (PYPLAS) is developed to assist students in learning Python programming and reduce the workload for university instructors. It provides fill-in-the-blank exercises ranging from beginner to advanced levels, generated using a blank element selection algorithm.
- According to a publication titled "Studying the Impact of Auto-Graders Giving Immediate Feedback in Programming Assignments" compiled by (Mitra, 2023), while students use immediate input from an auto-grader, it does not discourage them from developing independent testing skills. Furthermore, feedback enables students, particularly those from underrepresented groups (such as women), to learn more successfully and confidently.

RESEARCH METHODOLOGY

Time and Place: conducted at the State Polytechnic of Malang for 3 months, from April 2024 to July 2024.

Data Collection Method:

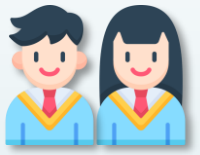
Responden: 40 Students of Information Technology Major at Politeknik Negeri Malang

Platform: - Google Colaboratory

- Google Drive

- Google Forms

VOLUNTEERS DATA AND TESTING



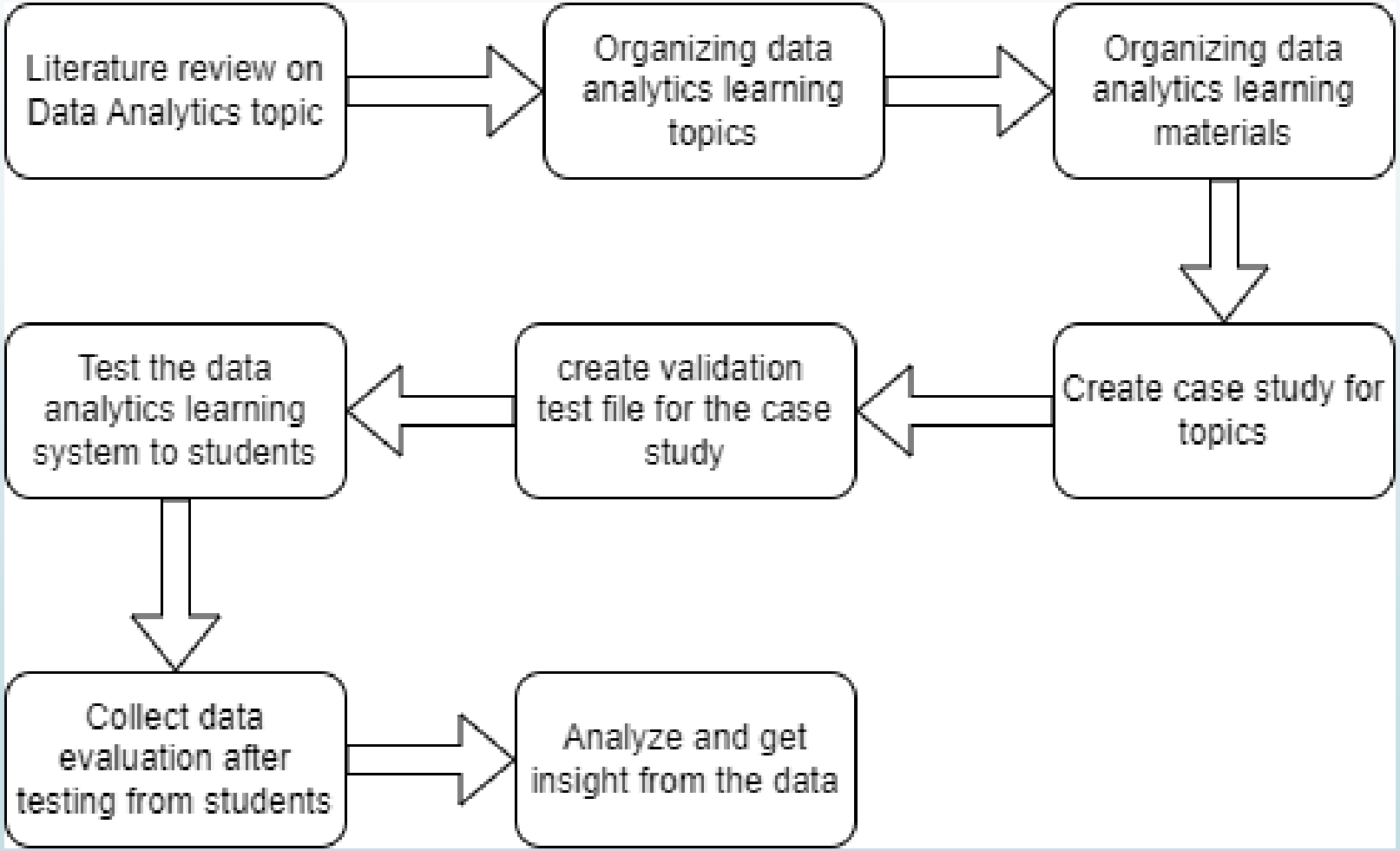
The total number of volunteers is **40 students** from the **Department of Information Technology** at Politeknik Negeri Malang.

Testing Devices : Laptop or PC

Testing Time : 3 days

System Design

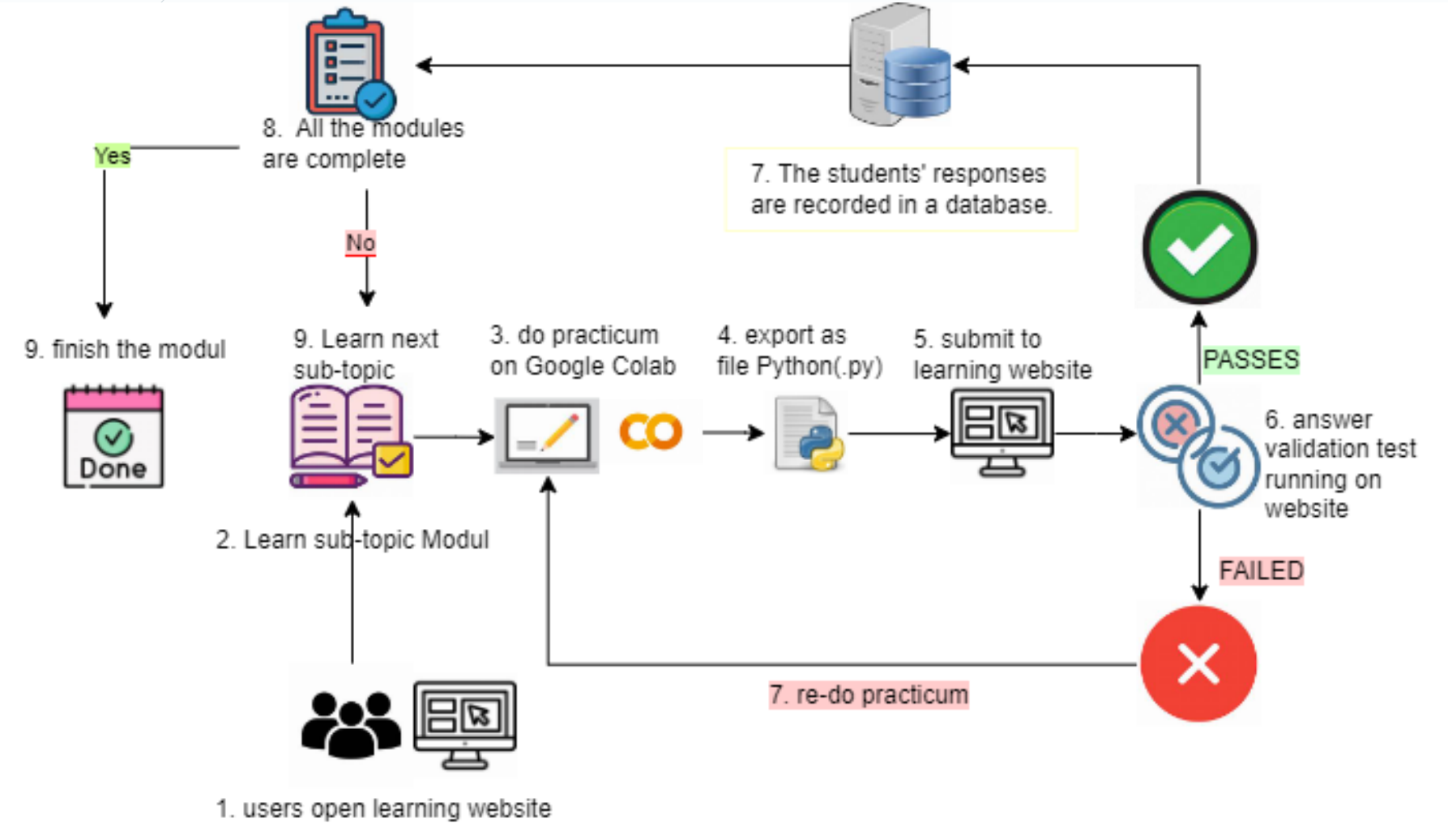
This research focuses on the application of Unit Testing using Codewars_test in the context of Data Analytics learning materials.



LEARNING MATERIALS

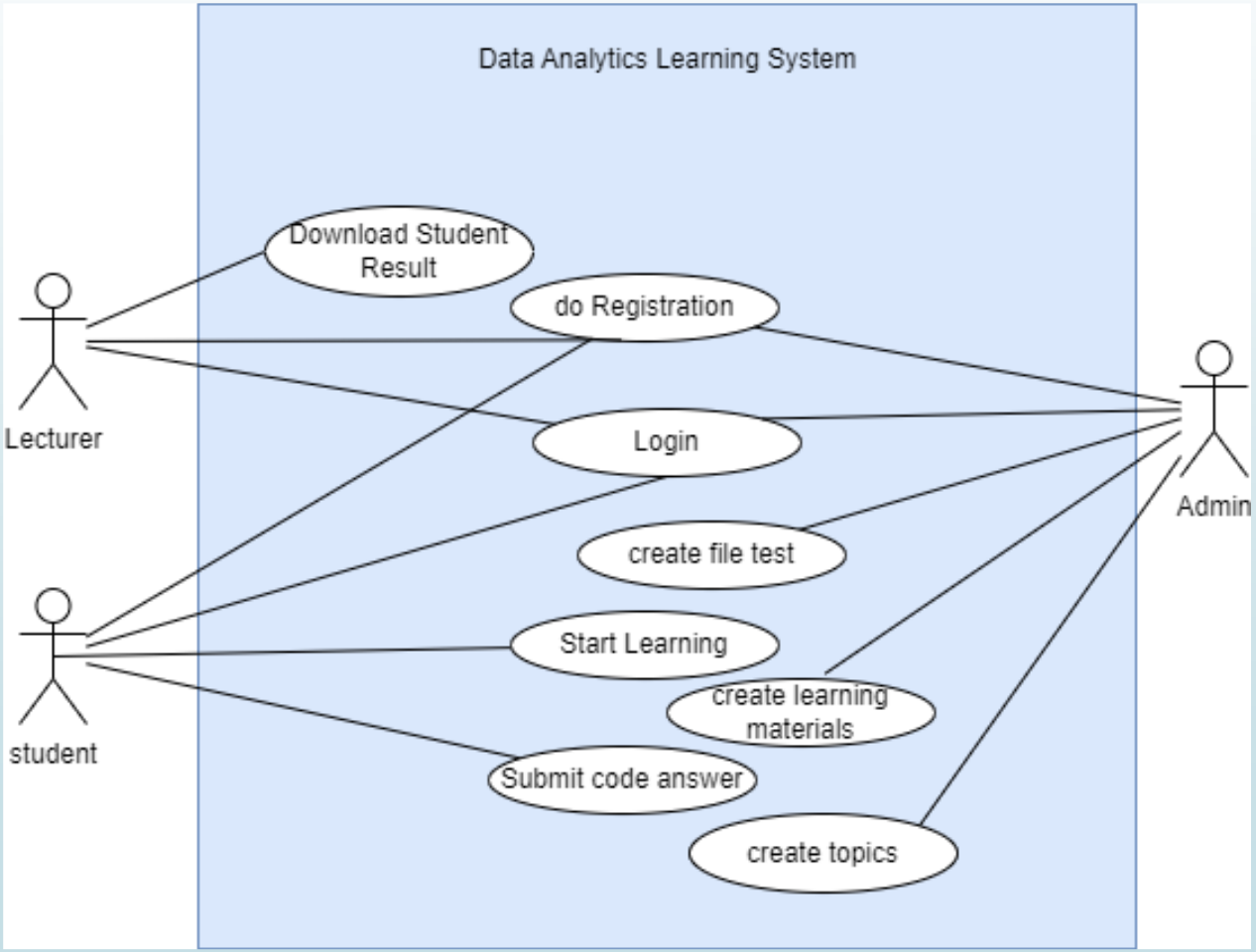
No	Experiment Name	Topic
1.	Get the dataset from url and read as CSV file.	Data Loading
2.	Display the Produk Online Store Data.	Data Loading
3.	Cleaning column name for real estate data.	Cleaning Dataset
4.	Handling missing values for Calories data.	Cleaning Dataset
5.	Cutting a NumPy 2D array on a specific coverage.	Data Manipulations
6.	Calculate Total Product Revenue.	Data Manipulations
7.	Selecting Specific Movies from Movie dataset.	Data Manipulations
8.	Analisa Measures of Variability pada customer's Age.	Data Manipulations
9.	Electronic Sales Analysis: Understanding Sales Trends Through Measures of Central Tendency.	Data Manipulations
10.	Visualizing the age distribution of the population based on age range using a Pie Chart.	Data Visualization

System Architecture Design



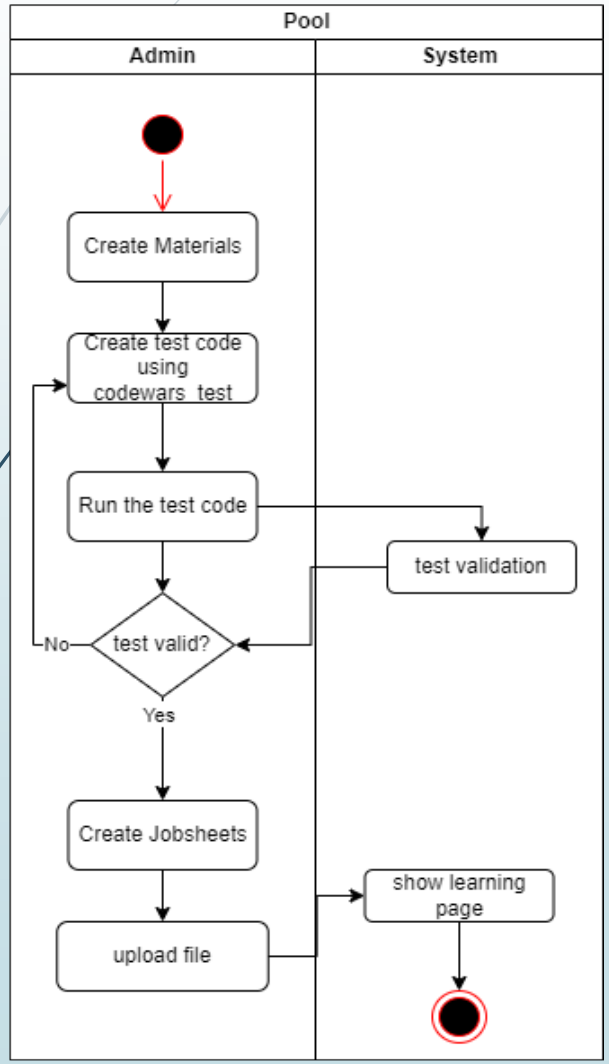
Use Case Diagram

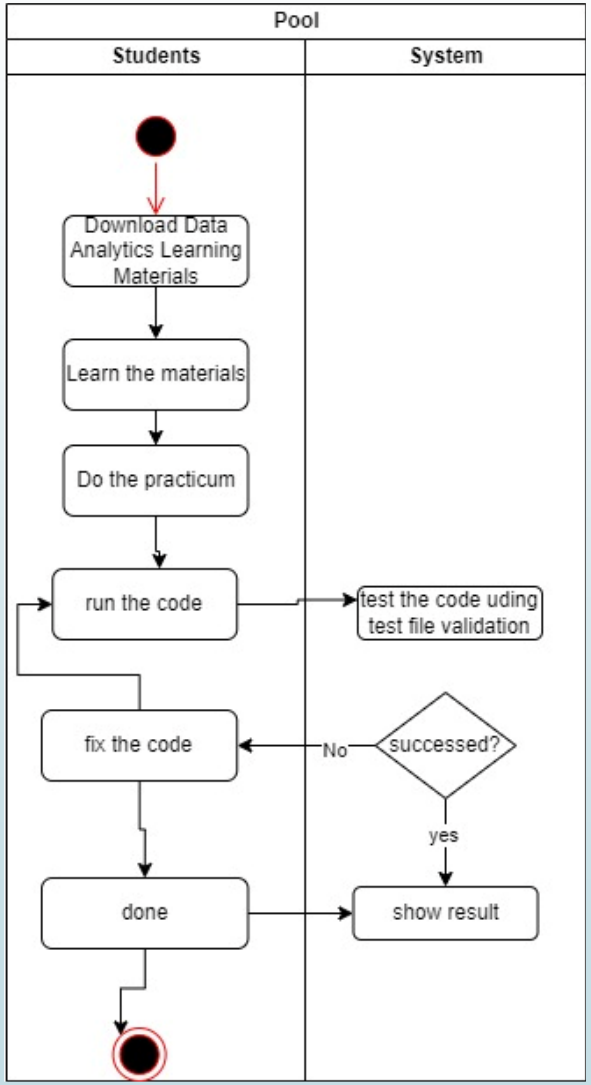
- the focus is limited to the development of use cases such as **creating a topic, creating learning materials, and create file test**.

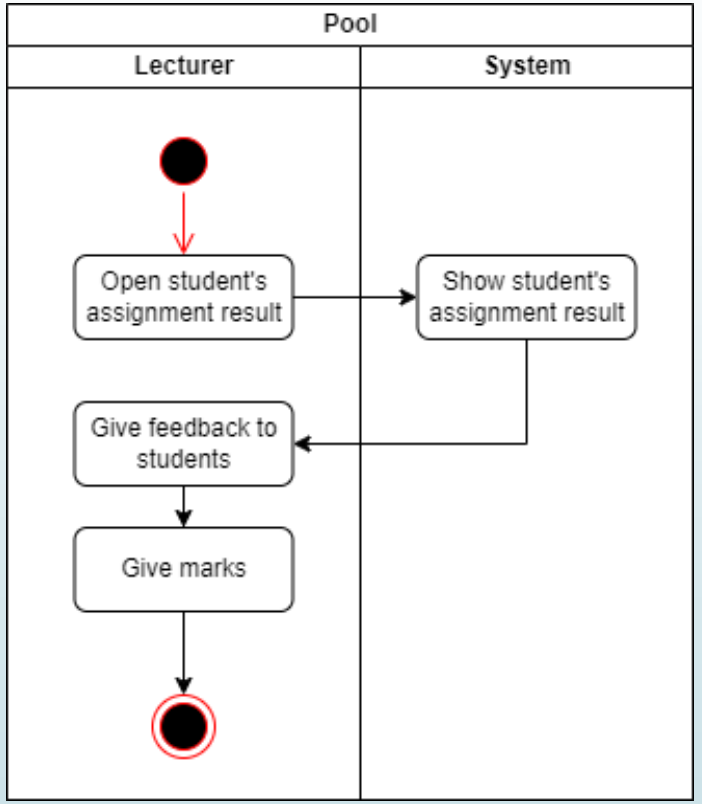


Activity Diagram

- **creating a topic, creating learning materials, and create file test.**







IMPLEMENTATION AND TESTING

- Limitations of Implementation

1. Development of tasks using Google Colaboratory.
2. Application testing is done using Codewars_test framework.

- **5.2 Learning Implementation**

1. Data Analytics Learning Modul
2. Test file using Codewars_test for validation

Modul Descriptions

No	Experiment Name	Topic	Description
1.	Get the dataset from url and read as CSV file.	Data Loading	Students were given an experiment to access data sets from external sources into the system. Using this function pd.read_csv(url)
2.	Display the Produk Online Store Data.	Data Loading	Students are given an experiment to delete data columns that are not needed in the data analysis process. Display 5 first rows in the dataset using head_rows().
3.	Cleaning column name for real estate data.	Cleaning Dataset	Students were given an experiment to tidy up the column name arrangement to make it uniform. Delete the column using df.rename(columns=clean_names).
4.	Handling missing values for Calories data.	Cleaning Dataset	Students are given a practicum to handle missing values in a column and fill them with the column's mean value using the fillna() function.

Modul Descriptions

5.	Cutting a NumPy 2D array on a specific coverage.	Data Manipulations	Taking product-specific sales data and applying slicing to extract specific data on the Numpy 3D array.
6.	Calculate Total Product Revenue.	Data Manipulations	Calculating Total Product Revenue using the sum() function.
7.	Selecting Specific Movies from Movie dataset.	Data Manipulations	Analyses film datasets to find 10 films using argsort() and iloc() functions.
8.	Analisa Measures of Variability pada customer's Age.	Data Manipulations	Analisa Measures of Variability pada customer's Age using mean(), median(), std(), skew(), quantile(), and corr() function.

Modul Descriptions

9.	**Electronic Sales Analysis: Understanding Sales Trends Through Measures of Central Tendency.**	**Data Manipulations**	**Electronic Sales Analysis: Understanding Sales Trends Through Measures of Central Tendency using mean(), median(), and mode() function.**
10.	**Visualizing the age distribution of the population based on age range using a Pie Chart.**	Data Visualization	analyze the age distribution of the population of a city by visualizing it using pie charts.

KODE TEST VALIDASI

```python
@codewars_test.describe("BAB 2 | Percobaan 2")
def fixed_tests():

    @codewars_test.it("1. Test Memuat Data")
    def test_load_data():
        print("============================================================================")
        # Assuming expected columns are "Customer ID", "Age" and "Total Spent (USD)"
        expected_columns = ["Invoice ID","Branch", "City", "Customer type", "Gender", "Product line", "Unit price", "Quantity", "Tax 5%", "Total", "Date", "Time", "Pay
        try:
            codewars_test.assert_equals(list(pc.data_load().columns), expected_columns, "====> URL dataset yang digunakan tidak sesuai; kolom pada dataset berbeda")
        except Exception as e:
            codewars_test.fail(f"====> Error loading data; Terdapat Typo pada kode: {str(e)}")

        expected_rows = 1000
        try:
            codewars_test.assert_equals(len(pc.data_load()), expected_rows, "====> Periksa kembali URL yang digunakan; jumlah data pada dataset berbeda")
        except Exception as e:
            codewars_test.fail(f"====> Error checking row count; Terdapat Typo pada Kode: {str(e)}")
```

Figure 5. 4 Loading Data Test

KODE TEST VALIDASI

```python
@codewars_test.it("2. Test Fungsi head_rows()")
def test_show_first_five_rows():
    print("================================================================================")
    # Call the function and capture the output
    try:
        actual_output = pc.head_rows().shape[0] # Get the number of rows

        expected_rows = 5

        # Assert that the expected output matches the actual output
        codewars_test.assert_equals(actual_output, expected_rows, "Should Showing First Five Rows" )
    except Exception as e:
        codewars_test.fail(f"====> Error pada head_rows(); Terdapat Typo pada kode function: {str(e)}")

    except KeyError as e:
                codewars_test.fail(f"====> Error: Kolom '{e.args[0]}' tidak ditemukan dalam DataFrame. Pastikan Anda memiliki kolom '{e.args[0]}' dalam dat
```

Figure 5. 5 Function test

```
@codewars_test.it("5. Test Print Fungsi Jumlah Pendapatan")
def test_total_revenue():
    print("=======================================================================")
    expected= "307587.38"
    output_lines = cmd.stdout.decode().splitlines()
    if output_lines:
        actual_value = output_lines[0]
    else:
        actual_value = " "

    codewars_test.assert_equals(actual_value, expected, '====> Error :Tidak Menampilkan nilai fungsi jumalah_pendapatan() menggunakan print()')
```

Figure 5. 6 Output Test

```
<DESCRIBE::>BAB 1 | Percobaan 1
==========================================================================

<IT::>1. Test Memuat Data

<PASSED::>Test Passed

<PASSED::>Test Passed

<COMPLETEDIN::>50.77

<IT::>2. Test Print Nilai Fungsi data_load()
--------------------------------------------------------------------------

<FAILED::>====> Error :Tidak Menampilkan nilai fungsi data_load() menggunakan print(): False should equal True

<COMPLETEDIN::>0.03

<COMPLETEDIN::>50.88
```

Validation Result Output

RESULTS AND DISCUSSION

40 students of the Faculty of Information Technology of the State Polytechnic of Malang as **the participant**

No	Name	NIM	Major
1	Amalia Nuraini	2041720160	Jurusan Teknologi Informasi
2	Elvira Sania Mufida	2041720080	Jurusan Teknologi Informasi
3	Rabiatul Fitra Aulia	2041720154	Jurusan Teknologi Informasi
4	Neha Viranica Nauly	2141764127	Jurusan Teknologi Informasi
5	Khofifah Amanda	2041720119	Jurusan Teknologi Informasi
6	Shine Devi Oktaviana Ronix Syah Putri	2041720065	Jurusan Teknologi Informasi
7	Lelyta salsabila	1941720026	Jurusan Teknologi Informasi
8	Salwa Zhafira Pratiwi Wahyudi	2041720138	Jurusan Teknologi Informasi
9	FAHREZA PRIMA HAKIM	2041720210	Jurusan Teknologi Informasi
10	Nanda Shabrina Putri Kurnia	2141762064	Jurusan Teknologi Informasi
11	Mohammad Izamul Fikri Fahmi	2141720171	Jurusan Teknologi Informasi
12	Annisa Aulia Nadhila	2041720023	Jurusan Teknologi Informasi
13	Yoby Ryaian Pratama	2041720039	Jurusan Teknologi Informasi
14	Maulana Bintang I.	2041720132	Jurusan Teknologi Informasi
15	Izzatun Nauly	2041720166	Jurusan Teknologi Informasi

RESULTS AND DISCUSSION

40 students of the Faculty of Information Technology of the State Polytechnic of Malang as **the participant**

16	**Deatrisya Mirela Harahap**	**2041720013**	**Jurusan Teknologi Informasi**
17	Iftitah Hidayati	2041720006	Jurusan Teknologi Informasi
18	Hafidz Irwan M	2141764079	Jurusan Teknologi Informasi
19	Ahmad Farrel Sirajudin Zaidan	2041720238	Jurusan Teknologi Informasi
20	Della Jannata Febiana	2041720034	Jurusan Teknologi Informasi
21	Eva Monika Septiana	2141764017	Jurusan Teknologi Informasi
22	Hilda Khoirotul Hidayah	2041720161	Jurusan Teknologi Informasi
23	Dherisma Hanindita Utami	2041720018	Jurusan Teknologi Informasi
24	Yaldika Putra Sinaga	2041720056	Jurusan Teknologi Informasi
25	Rifqi Alauddin Nur	2041720169	Jurusan Teknologi Informasi
26	farah zulfa hamidah	2041720069	Jurusan Teknologi Informasi
27	Andre Maulana Mustofa	2041720211	Jurusan Teknologi Informasi
28	Abdulilah Ali Qaid Al-shabany	2041720203	Jurusan Teknologi Informasi
29	Chan Paul Amol	2041720202	Jurusan Teknologi Informasi
30	Amern Mohammed Ali Albaiti	2041720204	Jurusan Teknologi Informasi

RESULTS AND DISCUSSION

40 students of the Faculty of Information Technology of the State Polytechnic of Malang as **the participant**

31	**Muhammad Hafiz Sayyid Diangga**	**2041720156**	**Jurusan Teknologi Informasi**
32	Atmayanti	2041720016	Jurusan Teknologi Informasi
33	Silvia Prada Aprilia	2041720141	Jurusan Teknologi Informasi
34	Muhammad Ilham El Hakim	2041720162	Jurusan Teknologi Informasi
35	Ibnu Khalis Rabbani	20415720159	Jurusan Teknologi Informasi
36	Zaed Abdullah	2141764148	Jurusan Teknologi Informasi
37	Akhmadheta Hafid Prasetyawan	2041720221	Jurusan Teknologi Informasi
38	Mochammad Hairullah	2041720074	Jurusan Teknologi Informasi
39	M. Thosin Yuhaililul Hilmi	2141764032	Jurusan Teknologi Informasi
40	Hanif Widyantoro	2141764039	Jurusan Teknologi Informasi

Student's Completion Time

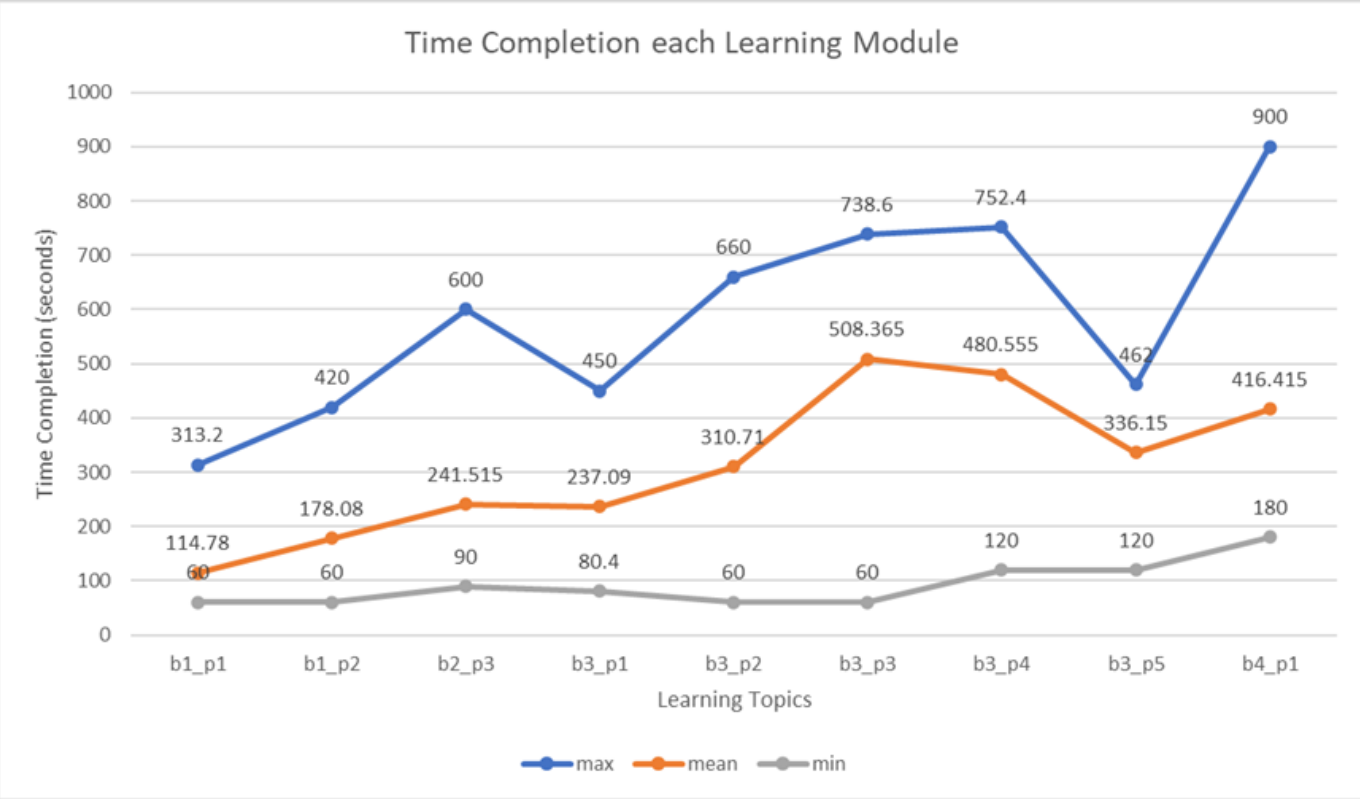


Table displays the outcomes of implementing the new teaching method on a group of 40 students. It is a graph **of the results of the test based on the student's work time of each learning topic**. The bottom **green line** indicates the fastest time a student has spent completing each learning topic, the **red line** shows the longest time the student has had spent finishing each learning subject, whereas the **blue line** between the **green** and **red** lines represents the average length of time it takes a student to complete each learning theme.

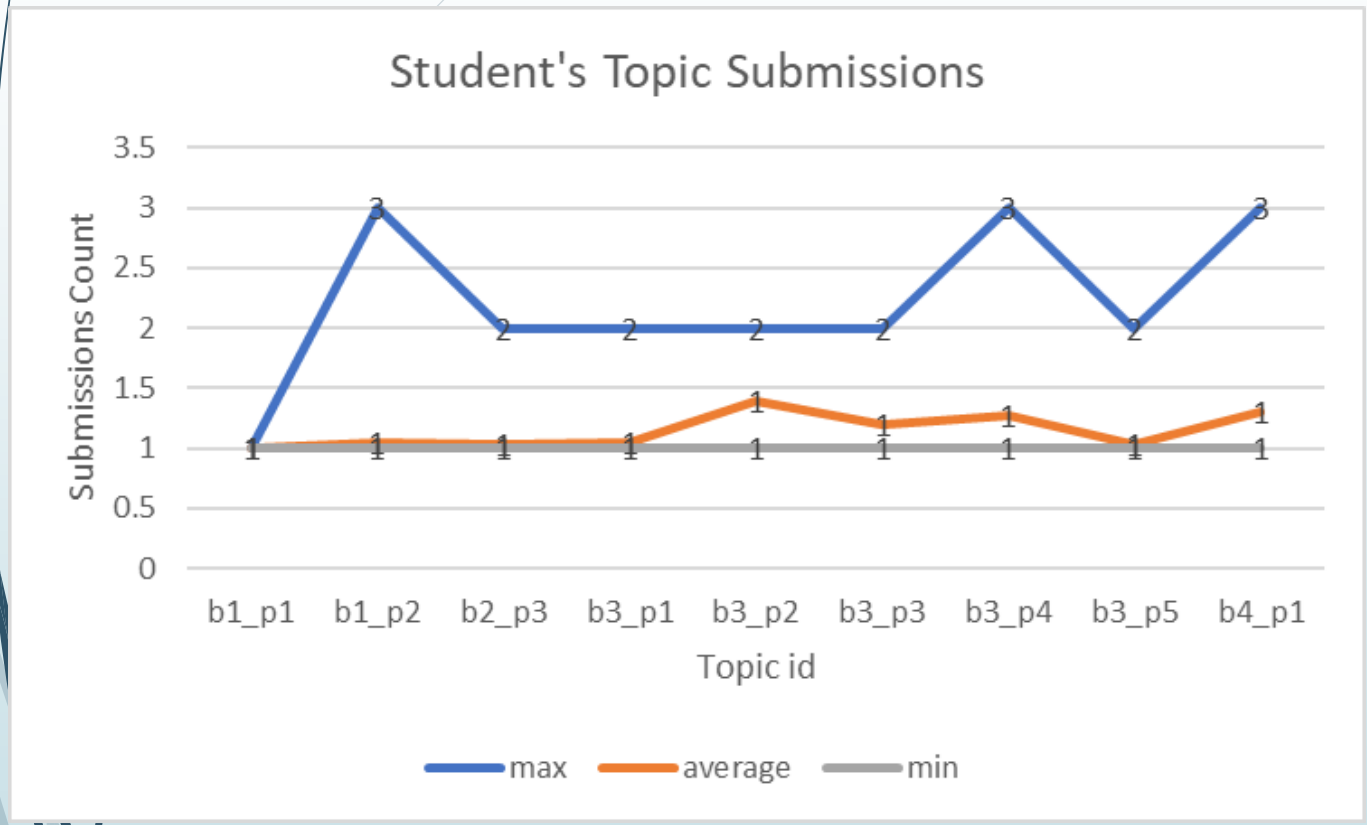
	b1_p1	b1_p2	b2_p2	b2_p3	b3_p1	b3_p2	b3_p3	b3_p4	b3_p5	b4_p1
max	840	900	900	900	900	840	900	900	900	900
mean	246	309	354	384	411	469	542	523	516	534
min	60	60	60	60	60	60	60	60	60	120

Student's Completion Time

Results showed that all students successfully completed a series of tasks, albeit at varying times. Based on the average completion time for each module, the data analytics learning modules were categorized as shown in the table.

Level	Learning Stages
Easy	b1p1,b1p2
Medium	b2p3, b3p1, b3p2, b3p5
Hard	b3p3, b3p4, b4p1

The difficulty level for each topic



From this figure, **the submission counts vary across the topics**. Some topics have received a high number of submissions, while others have received relatively few. maximum number of submissions for a single topic is 3. This indicates that **some students are submitting multiple times** for certain topics, which could be due to making revisions. The **average number of submissions per topic falls between 1 and 2**. This suggests that, on average, students are submitting 1-2 times per topic. The minimum number of submissions for a topic is 1. This means that there are no topics with zero submissions, and **each topic has received at least one submission**.

FEEDBACK MAHASISWA

The constructive suggestions gathered from student feedback are as follows:

- The consistency of the language used in the module to use one language or make two versions.
- The module should be more detailed. It doesn't matter if the sentences are long as long as they are detailed and easy to understand.
- Perhaps the module presentation could be made more creative to make it a bit more colorful.
- the guidance in the jobsheet needs to be clarified for better understanding.

Here are some examples of positive feedback we received from students:

- It's good, because it's possible to know if the answers are right or wrong through testing.
- the explanations and guides on the jobsheet make it very easy for the user to understand the material.
- The jobsheet allowed me to learn independently about data analytics.
- I hope it can be applied as a learning module"}.

Conclusion

- The implementation of test-driven development (TDD) using Codewars_test in a Python programming learning context, as demonstrated with the 40 student participants from the Department of Informatics Engineering at Politeknik Negeri Malang, has yielded positive results. The provided module materials were found to be effective in facilitating students' understanding of data analytics concepts and guiding them through practical tasks. The positive feedback and suggestions provided by the students further validate the effectiveness of the learning system. Students provided 85.71% positive feedback. Their comments indicate that the module materials were clear, easy to follow, and helpful in facilitating independent learning.
- The findings of this study suggest that the integration of TDD with Codewars_test in a data analytics learning assistance system can be a valuable approach for enhancing students' learning experiences and outcomes. By providing immediate feedback, encouraging a structured approach to problem-solving, and fostering independent learning, TDD can help students develop a deeper understanding of the underlying concepts and improve their coding skills.