

## Basic Introduction to NLP: Text and Language Analysis with Python

El Faath Sadam<sup>1</sup>, Ananda Radhiyya Annafi<sup>2</sup>, Iqlal Wira Prasetio<sup>3</sup>,  
Muhamad Sulton Nashiro<sup>4</sup>, Reihan Septyawan<sup>5</sup>, Mirza Sutrisno<sup>6</sup>,  
Rita Dewi Risanty<sup>7</sup>, Sitti Nurbaya Ambo<sup>8</sup>, Yana Adharani<sup>9</sup>, Rully Mujiastuti<sup>10</sup>

<sup>1,2,3,4,5,6,7,8,9,10</sup> Teknik Informatika, Universitas Muhammadiyah Jakarta

e-mail: [20210410700062@student.umj.ac.id](mailto:20210410700062@student.umj.ac.id)

### Abstrak

Natural Language Processing (NLP) telah menjadi bidang penting dalam komputasi modern, dengan fokus pada interaksi antara bahasa manusia dan komputer. Webinar dan workshop "Pengenalan Dasar NLP: Analisis Teks dan Bahasa dengan Python" dilaksanakan pada 18 Juli 2024 secara daring, dihadiri oleh 35 peserta. Kegiatan ini meliputi pengenalan tentang NLP, demonstrasi penggunaan Python untuk analisis teks, dan sesi praktik dengan Google Colab. Hasil pre-test menunjukkan bahwa 70,4% peserta memiliki pengetahuan awal yang benar, meningkat menjadi 81,8% pada post-test. Feedback dari 44 peserta menunjukkan tingkat kepuasan yang tinggi, 100% peserta setuju bahwa materi yang disampaikan sesuai dan terstruktur dengan baik, narasumber kompeten, dan kualitas teknis kegiatan baik. Workshop ini berhasil meningkatkan pemahaman peserta tentang NLP, teknik pra-pemrosesan data teks, dan penggunaan Python dalam analisis teks bahasa. Kesimpulannya, kegiatan ini efektif dalam memperkenalkan NLP dan meningkatkan keterampilan analisis teks peserta menggunakan Python.

**Kata kunci:** *Analisis Teks, Google Colab, Natural Language Processing (NLP), Pra Pemrosesan Data Teks, Python*

### Abstract

"Natural Language Processing (NLP) has emerged as a crucial field in modern computing, focusing on the interaction between human language and computers. On 18 July 2024, the online webinar and workshop titled "Basic Introduction to NLP: Text and Language Analysis with Python" was attended by 35 participants. The session included an introduction to NLP, a demonstration of using Python for text analysis, and a practical session with Google Colab. The pre-test results revealed that 70.4% of the participants had correct initial knowledge, which increased to 81.8% in the post-test. Feedback from 44 participants indicated a high level of satisfaction, with 100% of participants agreeing that the material presented was appropriate and well-structured, the speakers were competent, and the technical quality of the activities was good. The workshop successfully boosted participants' understanding of NLP, text data pre-processing techniques, and the use of Python in language text analysis. In conclusion, this workshop effectively introduced NLP and enhanced participants' text analysis skills using Python."

**Keywords:** *Google Colab, Natural Language Processing (NLP), Python, Text Analytics, Text Data Pre-processing*

### PENDAHULUAN

Kecerdasan buatan adalah cabang ilmu komputer yang berhubungan dengan otomasi perilaku yang cerdas (1993). Kemudian menurut Haag dan Peter kecerdasan buatan adalah bidang studi yang berhubungan dengan penangkapan, pemodelan, dan penyimpanan kecerdasan manusia ke dalam sebuah sistem teknologi informasi sehingga sistem tersebut dapat digunakan sebagai proses pengambilan keputusan yang dilakukan oleh manusia (1996). Dari dua pengertian tersebut dapat disimpulkan bahwa kecerdasan buatan adalah cabang ilmu komputer yang membahas tentang penangkapan, pemodelan, dan penyimpanan kecerdasan manusia ke dalam

sebuah teknologi informasi yang nantinya dapat dimanfaatkan untuk pengambilan keputusan (Luger, George F., & William A, 1993).

### **Natural Language Processing (NLP)**

Natural Language Processing (NLP) adalah bidang linguistik yang memungkinkan mesin memahami bahasa manusia. NLP menangani data teks yang kompleks dan tidak terstruktur, termasuk tugas-tugas seperti segmentasi kata, penandaan bagian ucapan, pengenalan entitas bernama, ekstraksi relasi, dan terjemahan mesin. Pembelajaran representasi membantu merepresentasikan data bahasa alami dalam berbagai tingkat dan domain, meningkatkan efektivitas dan ketahanan kinerja NLP. NLP menggunakan teori dan teknik untuk menganalisis konten tekstual, memungkinkan interaksi manusia-komputer melalui percakapan, di mana komputer mengenali dan memproses perintah serta merespons (Thomas & Devi, 2021; Noor Fariza, 2022).

### **Google Colab**

Google Colab atau Google Colaboratory adalah sebuah executable document yang dapat digunakan untuk menyimpan, menulis, serta membagikan program yang telah ditulis melalui Google Drive. Software ini pada dasarnya serupa dengan Jupyter Notebook gratis berbentuk cloud yang dijalankan menggunakan browser seperti Mozilla Firefox dan Google Chrome. Google Colab memungkinkan penggunaannya untuk menjalankan kode bahasa pemrograman python tanpa perlu melakukan proses instalasi dan setup. Semua keperluan setting dan adjustment akan diserahkan ke cloud (Oliver, 2022).

### **Masalah dan Rencana pemecahan Masalah**

Dalam webinar yang berjudul "Basic Introduction to NLP: Text and Language Analysis With Python", masalah yang akan dihadapi itu adalah bagaimana kita memperkenalkan konsep dasar dari Natural Language Processing (NLP) serta teknik-teknik analisis teks menggunakan bahasa pemrograman Python kepada peserta yang mungkin memiliki latar belakang yang beragam, mulai dari pemula hingga yang sudah memiliki pengalaman dasar dalam pemrograman. Rencana Pemecahan Masalah ini dapat dilihat sebagai berikut:

1. Menjelaskan gambaran umum tentang NLP: apa itu NLP, sejarah singkatnya, dan aplikasi praktis yang ada didalam kehidupan sehari-hari, misal seperti pencarian informasi dan chatbot
2. Menjelaskan tentang dasar-dasar pemrosesan teks seperti tokenisasi, stemming, lemmatization, stop words, dan part-of-speech tagging.
3. Pengenalan Python untuk NLP seperti NLTK, spaCy, dan gensim beserta membuat skrip Python sederhana untuk memulai analisis teks dasar.
4. Mengadakan Diskusi dan Tanya Jawab antara peserta dengan pemateri.

### **Tujuan**

Tujuan dari Webinar dan Workshop ini dilakukan yaitu:

1. Memberikan pemahaman tentang NLP ini beserta kegunaannya kepada para peserta.
2. Menjelaskan kepada para peserta tentang konsep dasar dan terminologi dalam NLP beserta dasar-dasar pemrosesan teks.
3. Memberikan pemahaman kepada peserta dalam menganalisis teks dan bahasa dengan Python menggunakan aplikasi website Google Colab

### **METODE**

#### **Edukasi melalui webinar**

Webinar dan workshop ini dilaksanakan pada Kamis, 18 Juli 2024, pukul 09.00 WIB -11.30 WIB secara daring melalui Zoom. Acara dimulai dengan pembukaan oleh MC yang menjelaskan tujuan dan materi acara. Tahap kedua adalah Pre-Test untuk mengukur pemahaman awal peserta tentang Natural Language Processing (NLP). Tahap ketiga menyampaikan materi dasar NLP yang mencakup pengertian, pentingnya dalam teknologi, contoh penggunaan, komponen utama, tokenisasi, normalisasi teks, stemming, lemmatization, dan analisis sentimen. Peserta juga diperkenalkan pada aplikasi AI dalam kehidupan sehari-hari. Tahap keempat adalah sesi tanya jawab dan Post-Test untuk mengevaluasi pemahaman peserta setelah mengikuti webinar.

### Pelatihan melalui Workshop

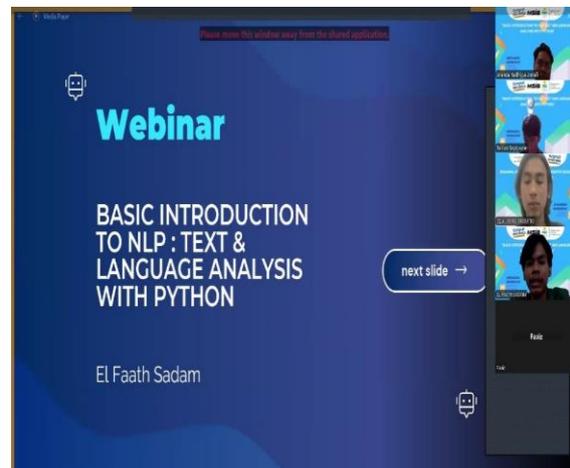
Dalam workshop ini, peserta akan memulai dengan mempersiapkan tools Google Colab dan membuka notebook baru sesuai petunjuk narasumber. Pada tahap pertama, narasumber akan memandu peserta dalam menggunakan Google Colab dan memperkenalkan source code sederhana untuk text preprocessing. Di tahap kedua, narasumber akan menjelaskan cara menganalisis teks dan bahasa menggunakan Python, termasuk teknik-teknik seperti tokenisasi, normalisasi teks, stemming, lemmatization, dan analisis sentimen. Pada tahap ketiga, sesi tanya jawab akan dibuka untuk memberikan kesempatan kepada peserta mengajukan pertanyaan dan mengatasi keraguan yang mungkin muncul selama penjelasan materi. Narasumber akan memberikan jawaban yang jelas dan mendalam, serta menjelaskan konsep-konsep tambahan jika diperlukan. Workshop ini bertujuan memberikan pemahaman praktis dan menyeluruh tentang text preprocessing dengan Python di Google Colab.

### Evaluasi melalui Feedback

Kami mengumpulkan feedback dari 44 peserta untuk mengevaluasi kegiatan webinar dan workshop. Hasilnya menunjukkan bahwa 100% peserta setuju materi sesuai dengan mata kuliah Teknik Informatika FTUMJ dan relevan dengan tema. Narasumber dinilai memiliki keahlian yang tepat dan mampu menjelaskan materi dengan baik. Semua peserta puas dengan kualitas audio dan visual selama sesi. Secara keseluruhan, 100% peserta menyatakan kepuasan terhadap acara ini. Evaluasi ini menunjukkan bahwa webinar dan workshop berhasil memenuhi harapan peserta dari segi materi, kualitas narasumber, dan aspek teknis.

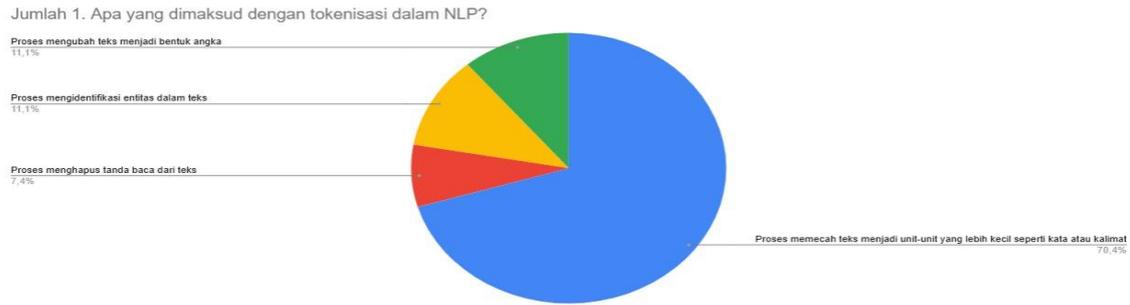
### HASIL DAN PEMBAHASAN

Pada diseminasi hasil MSIB Batch 6, webinar dan workshop dihadiri oleh 35 peserta. Acara ini dilaksanakan pada 18 Juli 2024 pukul 09.00-11.30 WIB melalui Zoom. MC membuka acara dengan menyampaikan tujuan dan materi yang akan dibahas. Setelah pembukaan dan menyanyikan lagu Indonesia Raya, materi "Basic Introduction to NLP: Text and Language Analysis With Python" disampaikan, mencakup pengenalan dasar NLP, pentingnya dalam teknologi, contoh penggunaan, komponen utama, serta teknik seperti tokenisasi, normalisasi teks, stemming, lemmatization, dan analisis sentimen.



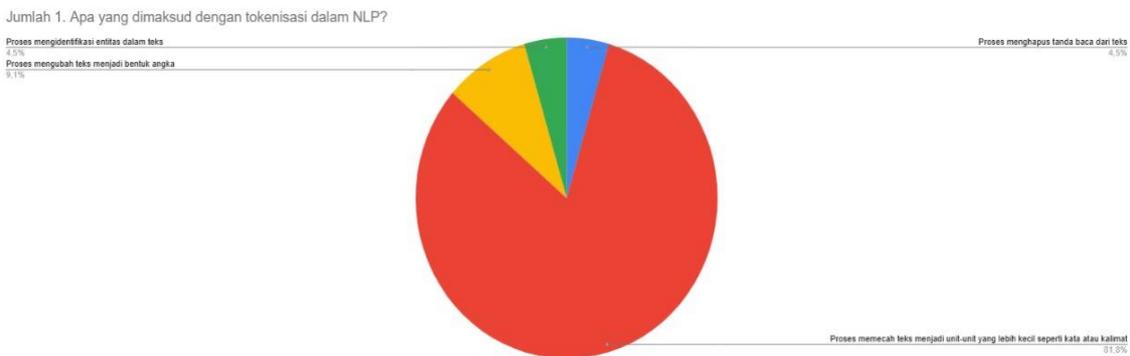
Gambar 1 dan 2 Pembukaan acara dan pemaparan materi webinar

Setelah webinar, sesi tanya jawab diadakan untuk menjawab pertanyaan peserta. Peserta kemudian diminta mengisi kuesioner melalui Google Form untuk evaluasi kegiatan, serta Form Pre-Test dan Post-Test. Berikut adalah lampiran persentase jawaban benar dari Pre-Test dan Post-Test:



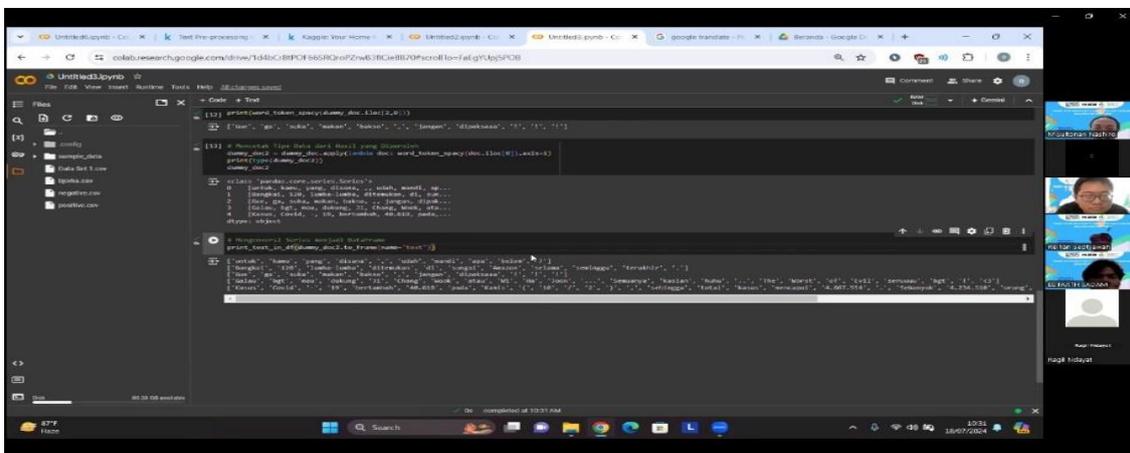
Gambar 2. Hasil Pre-Test No. 1

Hasil Pre-Test menunjukkan 70,4% peserta menjawab benar bahwa "Proses memecah teks menjadi unit-unit yang lebih kecil seperti kata atau kalimat." Persentase lainnya adalah 11,1% untuk "Proses mengubah teks menjadi bentuk angka," 11,1% untuk "Proses mengidentifikasi entitas dalam teks," dan 7,4% untuk "Proses menghapus tanda baca dari teks."



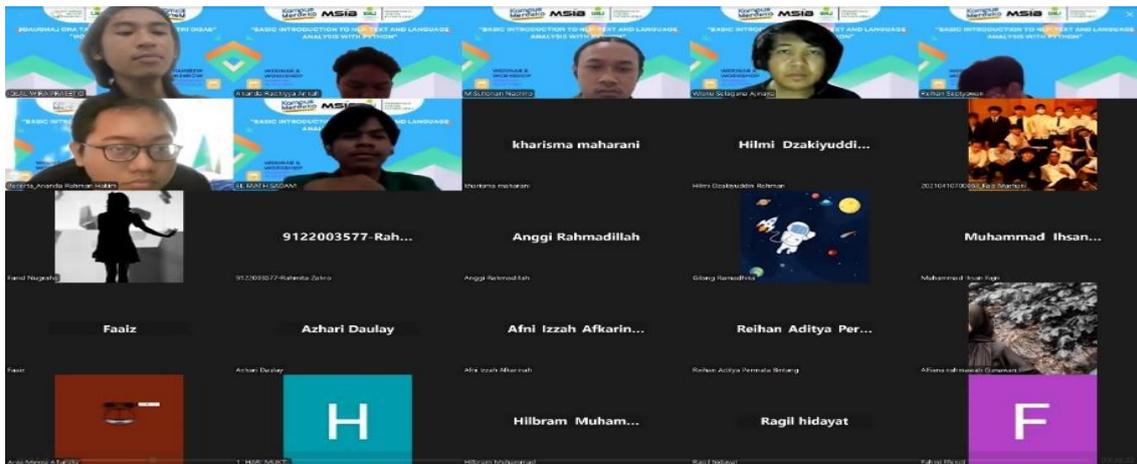
Gambar 3. Hasil Post-Test No. 1

Setelah Pre-Test, hasil Post-Test menunjukkan peningkatan pemahaman peserta dengan 81,8% menjawab benar "Proses memecah teks menjadi unit-unit yang lebih kecil seperti kata atau kalimat," dibandingkan 70,4% sebelum webinar. Persentase jawaban lainnya adalah 9,1% untuk "Proses mengubah teks menjadi bentuk angka," 4,5% untuk "Proses mengidentifikasi entitas dalam teks," dan 4,5% untuk "Proses menghapus tanda baca dari teks." Hasil ini menunjukkan bahwa webinar dan workshop berhasil meningkatkan pemahaman peserta dan materi yang disampaikan mudah dipahami. Kegiatan selanjutnya yaitu workshop dengan menggunakan tools Google Colab secara langsung untuk menganalisis teks dan bahasa, seperti yang ditunjukkan pada gambar 5.



Gambar 4. Penggunaan Tools Google Colab dari Workshop

Dalam workshop ini, kita akan menggunakan data science untuk analisis teks dan bahasa dengan Python di Google Colab. Kita akan memahami dan menerapkan text preprocessing untuk mengubah teks tidak terstruktur menjadi data terstruktur. Langkah pertama adalah mengumpulkan teks dari berbagai sumber, lalu melakukan preprocessing yang mencakup tokenisasi, penghapusan stop words, normalisasi teks, stemming, lemmatization, dan analisis sentimen. Kita akan menyiapkan lingkungan kerja di Google Colab dengan menginstal pustaka seperti NLTK, spaCy, TextBlob, dan VADER, lalu mengimpor pustaka tersebut untuk analisis teks. Selanjutnya, kita akan melakukan tokenisasi, menghapus stop words, dan melakukan normalisasi teks. Proses stemming dan lemmatization akan menyederhanakan kata-kata menjadi bentuk dasar. Terakhir, kita akan melakukan analisis sentimen untuk memahami emosi dalam teks. Peserta akan mendapatkan pemahaman komprehensif tentang teknik dasar analisis teks dan bahasa menggunakan Python. Setelah pemaparan materi, sesi tanya jawab akan diadakan untuk menjawab pertanyaan peserta.



Gambar 5 Sesi Tanya Jawab

Setelah sesi tanya jawab selesai, peserta diminta mengisi feedback yang disediakan oleh panitia melalui Google Form. Evaluasi ini bertujuan untuk mengetahui pengalaman peserta selama webinar dan workshop.

Pertanyaan kuesioner Feedback yang diajukan adalah sebagai berikut:

- 1) Apakah materi Webinar dan Workshop sesuai dengan mata kuliah yang diberikan di Teknik Informatika FTUMJ?
- 2) Apakah narasumber Webinar yang memberikan materi Webinar sesuai dengan bidang keilmuannya?
- 3) Apakah narasumber Workshop yang memberikan materi Workshop sesuai dengan bidang keilmuannya?
- 4) Apakah narasumber Webinar mampu menjelaskan materi dengan baik?
- 5) Apakah narasumber Workshop mampu menjelaskan materi dengan baik?

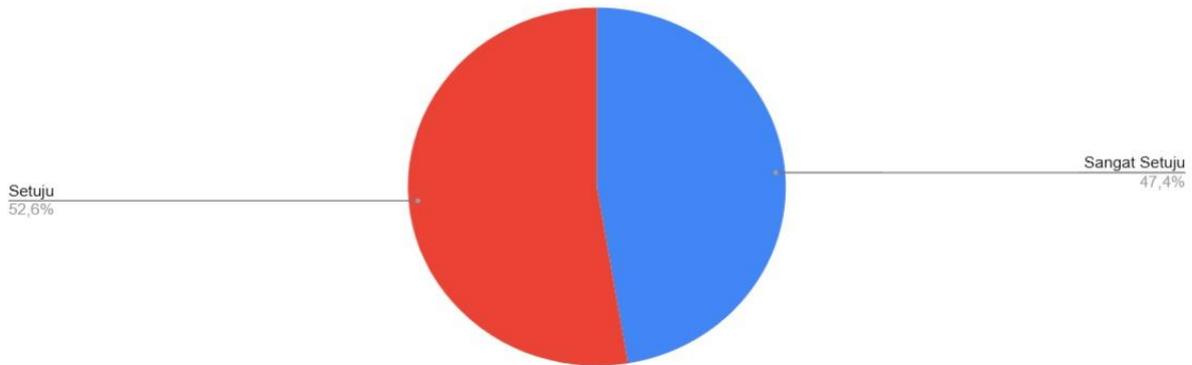
Selanjutnya, penilaian ini dilakukan menggunakan skala angka yang memiliki keterangan sebagai berikut:

- 1 = Sangat Tidak Setuju
- 2 = Tidak Setuju
- 3 = Netral
- 4 = Setuju
- 5 = Sangat Setuju

Dari hasil kuesioner yang telah diberikan dan diisi oleh 44 peserta, didapatkan hasil sebagai berikut:

1. Peserta menyatakan bahwa pemateri menyampaikan materi secara terstruktur dengan jumlah 47,4% sangat setuju, 52,6% setuju, 0% netral, dan 0% tidak setuju. Sebagaimana pada gambar 7 berikut.

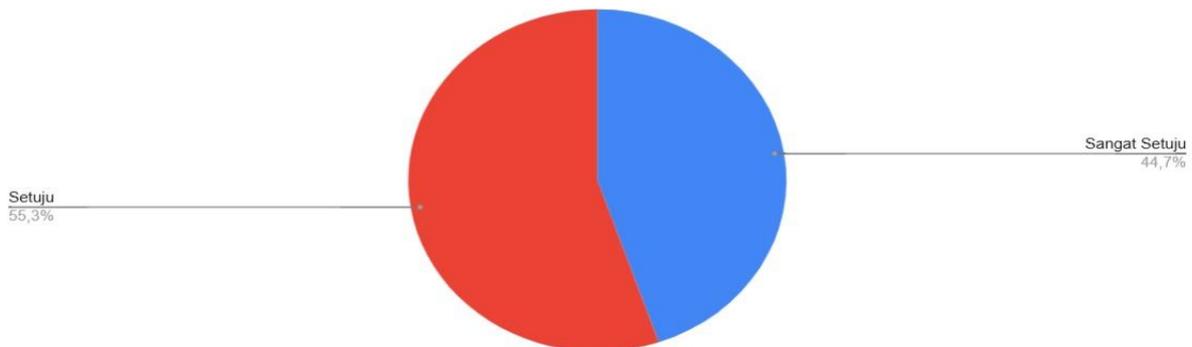
Jumlah Feedback [Apakah materi Webinar dan Workshop sesuai dengan mata kuliah yang diberikan di Teknik Informatika FTUMJ?]



**Gambar 6. Feedback**

2. Peserta menyatakan bahwa materi yang disampaikan sesuai dengan tema yang dibuat dengan jumlah 44,7% sangat setuju, 56,3% setuju, 0% netral, dan 0% tidak setuju. Sebagaimana pada gambar 8 berikut.

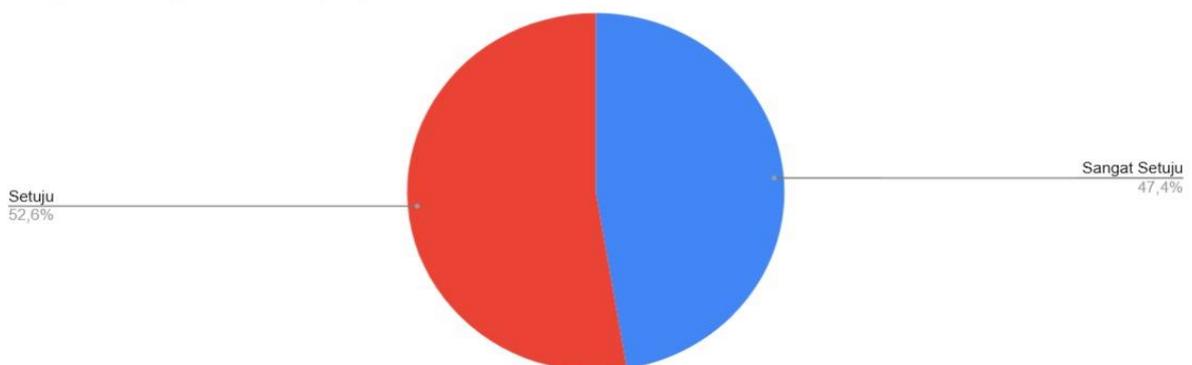
Jumlah Feedback [Apakah narasumber Webinar yang memberikan materi Webinar sesuai dengan bidang keilmuannya?]



**Gambar 7. Feedback 2**

3. Peserta menyatakan bahwa informasi yang disampaikan jelas dan tepat dengan jumlah 47,4% sangat setuju, 52,6% setuju, 0% netral, dan 0% tidak setuju. Sebagaimana pada gambar 9 berikut.

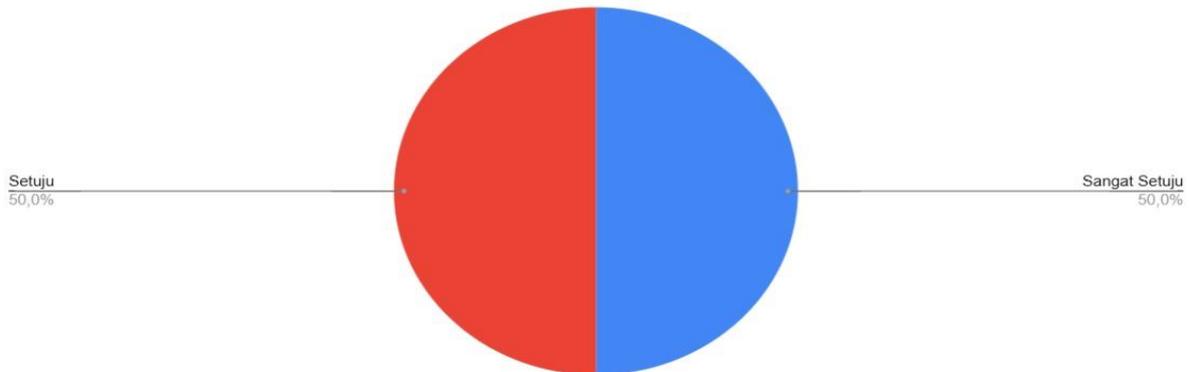
Jumlah Feedback [Apakah narasumber Workshop yang memberikan materi Workshop sesuai dengan bidang keilmuannya?]



**Gambar 8. Feedback 3**

4. Peserta menyatakan bahwa audio dan visual berjalan dengan baik selama sesi berlangsung dengan jumlah 50,0% sangat setuju, 50,0% setuju, dan 0% netral. Sebagaimana pada gambar 10 berikut.

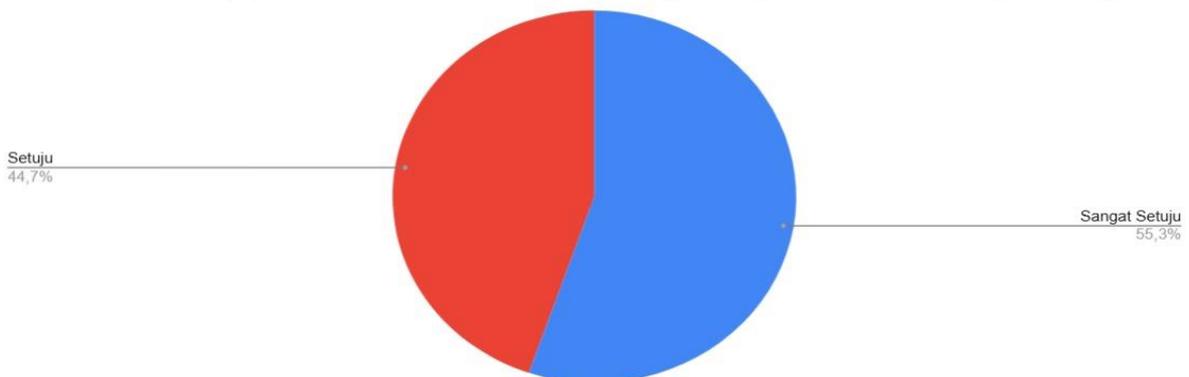
Jumlah Feedback [Apakah narasumber Webinar mampu menjelaskan materi dengan baik?]



**Gambar 9. Feedback 4**

5. Peserta menyatakan bahwa kepuasan terhadap keseluruhan kegiatan dengan jumlah sangat setuju dan setuju sama yaitu, sangat setuju 55,3%, setuju 44,7%, netral 0%, dan tidak setuju 0%. Sebagaimana pada gambar 11 berikut.

Jumlah Feedback [Apakah narasumber Workshop mampu menjelaskan materi dengan baik?]



**Gambar 10. Feedback 5**

Grafik diatas memperlihatkan bahwa narasumber dari webinar dan workshop ini sangat baik dalam memaparkan, menjelaskan, serta mengimplementasikan ke dalam aplikasi website *Google Colab*

## SIMPULAN

Webinar dan Workshop "Basic Introduction to NLP: Text and Language Analysis With Python" yang diikuti oleh 35 peserta berhasil meningkatkan pemahaman mereka tentang NLP dan analisis teks. Hasil Pre-Test menunjukkan bahwa 70,4% peserta sudah memiliki pengetahuan awal yang benar, sedangkan Post-Test menunjukkan peningkatan menjadi 81,8% jawaban benar, mencerminkan pemahaman yang lebih baik setelah acara. Dari 44 peserta yang mengisi kuesioner feedback, hasil evaluasi menunjukkan bahwa materi webinar sesuai dengan mata kuliah Teknik Informatika FTUMJ, dengan 47,4% sangat setuju dan 52,6% setuju. Narasumber webinar dan workshop dinilai sesuai dengan bidang keilmuannya, masing-masing dengan 44,7% dan 47,4% sangat setuju. Narasumber juga dianggap berhasil menjelaskan materi dengan baik, dengan 50% sangat setuju dan 50% setuju untuk webinar, serta 55,3% sangat setuju dan 44,7% setuju untuk

workshop. Peserta menyatakan kepuasan tinggi terhadap penyampaian materi dan kualitas pelaksanaan, dengan 55,3% sangat setuju dan 44,7% setuju terhadap kepuasan keseluruhan.

#### DAFTAR PUSTAKA

- Ali, A.-R. (2017), 'Pengenalan Natural Language Toolkit (NLTK) '.URL:[https:// code.tutsplus .com/id/tutorials/introducing-the-natural-language-toolkit-nltk-cms-28620](https://code.tutsplus.com/id/tutorials/introducing-the-natural-language-toolkit-nltk-cms-28620).
- Andana, M. H., Daffa, M., Fitria, N. U., & Mujiastuti, R. (n.d.). *Prosiding Seminar Nasional LPPM UMJ Website: <http://jurnal.umj.ac.id/index.php/semnaskat> E-ISSN: 2714-6286 Webinar & Workshop Natural Language Processing In The Life Of Artificial Intelligence. <http://jurnal.umj.ac.id/index.php/semnaskat>*
- Fariza, A. N. (2022, Januari 19). Mengenal Apa Itu Natural Language Processing (NLP). Retrieved November 20, 2022, from [Sekawanstudio.com](http://Sekawanstudio.com):
- Luger, George F., and William A. Stubblefield.1993.Artificial Intelligence Structures and Strategies for ComplexmProblem Soving 2nd edition.California: The Benjamin/Cumming Publishing Company Inc.
- Mirna Adriani, Jelita Asian, Bobby Nazief, Seyed MM Tahaghoghi, and Hugh E Williams. Stemming indonesian: A confix-stripping approach. *ACM Transactions on Asian Language Information Processing (TALIP)*, 6(4):1–33,2007.
- Oliver, A. (2022, Januari 25). Mengenal Google Colab: Mulai dari Definisi, Cara Menggunakan, hingga Manfaatnya. Retrieved November 20, 2022, from [glints.com](https://glints.com): <https://glints.com/id/lowongan/google-colab-adalah/#.Y-AqQXZBzI>
- Thomas, C. G., & Devi, J. (2021). A study and overview of the mobile app development industry. *International Journal of Applied Engineering and Management Letters (IJAEML)*, 5(1), 115-130.