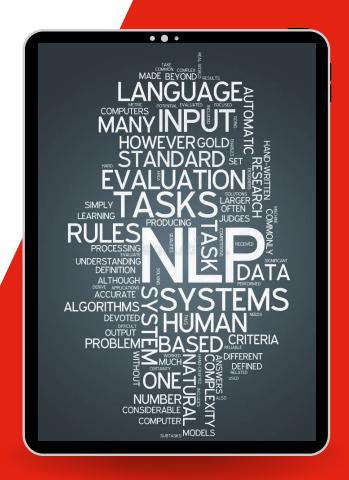
Introduction to NLP

Muhajir Akbar Hasibuan

Class Meeting 1

2023



Muhajir Akbar Hasibuan Data Scientist



■ muhajirakbarhsb@gmail.com

+62 85297023234

🎗 Jakarta, Indonesia

in linkedin.com/in/muhajirakbarhsb

github.com/muhajirakbarhsb

M medium.com/@muhajirakbarhsb

Experience

Telkom Indonesia

Sept 2020 - present

- Develop internal and external use cases
- · Provide data understanding in making a model
- · Provide Preparation and data engineering according to the use case implemented
- · Provide data validation so that the analysis results are as expected
- · Building modeling for the development or improvement of internal and external programs
- · Provide descriptive and diagnostic insight into data processing
- · Recommend and define new growth hacking strategy for digital marketing team

Project

- · Pioneered a Robust Big Data Solution for MyIndihome, Revolutionizing Customer Experiences
- Orchestrated a High-Performing ML Team, Elevating myIndihome TV's Personalized Content Impact by 25%
- Envisioned and Executed a Cutting-Edge Big Data Solution for Elevated Customer Engagement on mvIndihomeTV
- Fueled Business Insights via Dynamic Data Profiling, Performance Dashboards, and Insights for Langit Musik, RBT, and Upoint
- · Engineered Innovative Big Data Solutions that Propelled Growth for PadiUmkm (E-Commerce)
- · Powered Success for GameQoo through Strategic Big Data Solutions
- Architected and Established MLOps Framework, Elevating Telkom Indonesia's Digital Business Products
- Crafted Visionary Video Analytics Solutions for Telkom Indonesia's Revolutionary Digital Business IoT Product

MEMBER OF DATA SCIENTIST TASK FORCE | NOVEMBER 2021 - PRESENT MEMBER OF AI TASK FORCE | NOVEMBER 2022 - PRESENT

A Pool of data scientists and AI Engineer Expert in Telkom Indonesia. It was established to leverage data-driven culture for decision-making within the organization. (Applied Research, Standardization, Consultation)

Achievement

BEST Talent of the Year at Telkom Indonesia, Digital Business and Technology Division - Digital Technology and Platform -2022

Codex by Telkom Indonesia

- Building Data Pipeline for Langit Musik Recommender System
- · Business Analytic for Langit Musik
- · Build Recommender System for Langit Musik

Universitas Syiah Kuala

· Bachelor of Science, Statistics

Skills

Hard Skills

- · Data Analytics
- Statistics
- · Machine Learning
- · Deep Learning
- MlOns
- Business Intelligence
- · Data Engineering
- Cloud Computing
- Time Series & Demand Forecasting
- · Natural Language Processing
- Fraud & Anomaly Detection
- Computer Vision

- Tools Skills
- Python
- Pyspark
- Sql
- · Apache Airflow
- GCP
- Docker
 MLflow
- · Prometheus
- Evidently
- Grafana
- · Redash, Metabase, Superset, Looker Studio
- Pytorch

Apr 2020 - Sept 2020

2015-2020

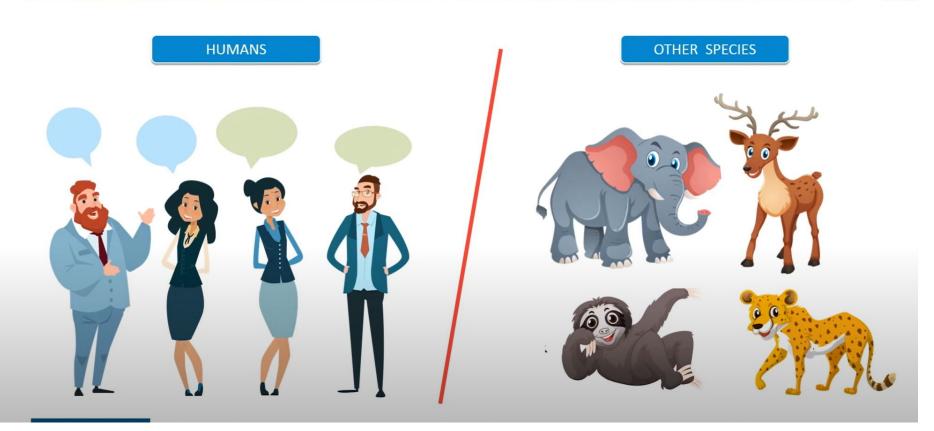
Class Meeting 1: Introduction to NLP (1 session) Introduction to NLP Basics of text data and its characteristics The importance of NLP in today's world Overview of the course structure and objectives	Class Meeting 7: Introduction to Topic Modeling(1 session) Introduction to topic Modelingg Topic Modeling Implementation LDA(Latent dirichlet allocation) Hands-on exercises with Topic Modeling	Class Meeting 14-15: Project Work (2 sessions) Dedicated sessions for students to work on NLP projects with guidance and assistance. Class Meeting 16: Project Presentations and Conclusion (1 session) Students present their NLP projects Recap of key takeaways from the course
Class Meeting 2: Text Preprocessing (1 session) Understanding the text preprocessing pipeline Tokenization, stemming, and lemmatization Stop words removal Hands-on exercises with Python for text preprocessing Class Meeting 3: Text Representation (1 session) Introduction to text representation techniques Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) n-gram Word embeddings (Word2Vec, GloVe)**just an introduction Practical exercises on text representation	Class Meeting 8: Mid-term Exam Class Meeting 9: Introduction to Word Embeddings (1 session) Fundamentals of word embeddings Word2Vec, GloVe, and FastText Word embedding applications Hands-on exercises with word embedding Class Meeting 10: Attention Mechanisms (1 session) Introduction to attention mechanisms Self-attention and multi-head attention Transformers architecture	Discuss further resources for NLP enthusiasts Course conclusion and feedback
Class Meeting 4: Introduction to ML, DL Supervised and Unsupervised (1 session) Introduction to deep learning (ML) Introduction to deep learning (DL) DL vs. traditional machine learning (ML) Machine Learning and their applications Neural networks and their applications Basic ML DL concepts and terminology	Practical examples of attention in NLP Class Meeting 11: Transformer Models (1 session) In-depth study of the Transformer model Pre-trained transformer models (BERT, GPT-2) Fine-tuning transformers for NLP tasks Transformer-based applications	
Class Meeting 5: Sentiment Analysis (1 session) What is sentiment analysis? Data collection and labeling for sentiment analysis Building a sentiment analysis model Practical sentiment analysis examples	Class Meeting 12: Advanced NLP Topics (1 session) Advanced NLP topics such as BERT, XLNet, and RoBERTa Transfer learning in NLP Ethics in NLP Recent developments and trends in NLP	
Class Meeting 6: Text Classification (1 session) Introduction to text classification Binary and multi-class classification Building a text classification model Real-world text classification examples	Class Meeting 13: Advanced NLP Topics and Deployment Process session) Introduction to LLM Introduction how industry utilize NLP to generate revenue Introduction MLOps for NLP(bonuses from practitioners)	in Industry (1

Agenda

- Overview of the course structure and objectives
- Introduction to NLP
- Basics of text data and its characteristics
- The importance of NLP in today's world

Introduction to Natural Language Processing







LANGUAGE



LANGUAGE



ALPHABETS 字母 वर्णमाला

ALFABETOS الأبجدية எழுத்துக்கள்



LANGUAGE



ALPHABETS

ং 字母 वर्णमाला ALFABETOS الأبجدية எழுத்துக்கள்



This is a Sentence





LANGUAGE



ALPHABETS

ং 字母 वर्णमाला ALFABETOS الأبجدية எழுத்துக்கள்



This is a Sentence



NLP?



Neuro Linguistic Programming

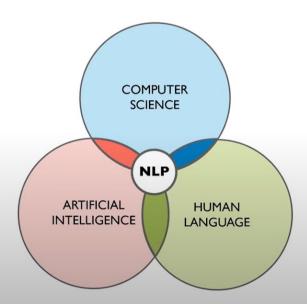
Natural Language Processing

What is NLP?



NLP: Natural Language Processing is a part of computer science and artificial intelligence which deals with human languages.





What is NLP?

The *GOAL* here is to process or understand Natural Language in order to perform useful tasks like

- Making Appointments
- Buying Things
- Spell Checking
- Generating Response
- Social Media Monitoring



Applications of NLP



Spell Checking

Information Extraction





Keyword Search Advertisement Matching



Applications of NLP



Sentimental Analysis

Chatbot

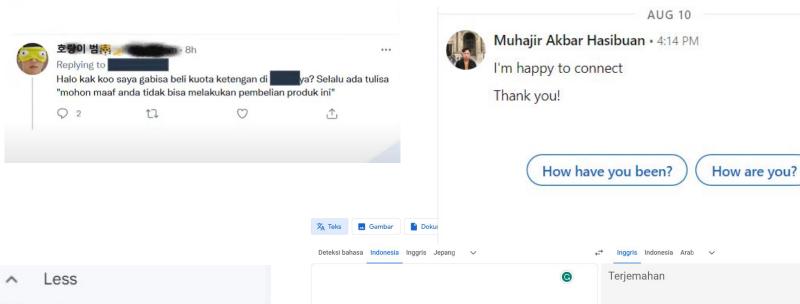


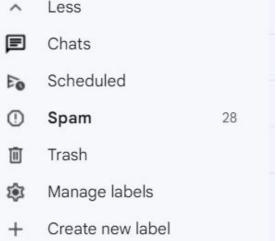


Speech Recognition

Machine Translation









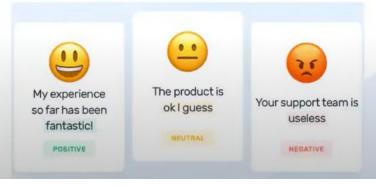
0 / 5.000

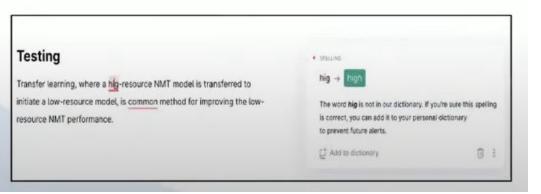
= -



TKDN 58% PKP











Structure VS Unstructure Data

Structure vs Unstructure

Key differences between structured and unstructured data

	Structured data	Unstructured data
ANALYSIS	Quantitative	Qualitative
SCHEMA CREATION	Schema-on-write	Schema-on-read
SEARCHING	Easy using SQL-based methods	May need special tools
FORMAT	Predefined, using alphanumeric characters	Typically non-character-oriented digital representations
STORAGE	May require more storage to accommodate defined data structures	Some forms require less storage; others have large file formats, requiring more storage
STORAGE FORMAT	Relational database management systems, data warehouses	Applications, NoSQL databases, data lakes

Structure vs Unstructure

```
] dfItem.head(30)['content'].values
  array(['Good',
          'Kalau mau lihat perangkat yang terhubung gak bisa, langsung auto keluar dari aplikasinya, tolong perbaiki lagi. 🙏 ',
          'Mohon di perbaiki bug aplikasi ini, banyak sekali bug, apalagi pada saat isi paket untuk bulan selanjutnya sering gagal padahal pulsa terisi dan mohon di tambah fitur auto perpanjang pada tgl yg sudah di tentukan ma
   tenggang paket habis dan bila ada pulsa yg ada di aplikasi.',
          TOLONG DI PERBAIKI YA KA APPLIKASI NYA SAAT SAYA PENCET PERANGKAT TERHUBUNG KELUAR SENDIRI MULU DARI APPLIKASI NYA 🔞 🔯 🔯 DAN AKUN YANG SUDAH ADA ENGGAK BISA KELUAR LAGI DARI APPLIKASI NYA KA DAN AKUN SAYA SUDAH K
  KALI LOGIN SUDAH KA KARNA AKUN NYA SERING KELUAR DARI APPLIKASI NYA DAN SEKARANG DI SURUH LOGIN LAGI AKUNNYA 🌡 🚳 '.
          'TIDAK DIREKOMENDASIKAN SAMA SEKALI., SUDAH BANYAK GANGGUANNYA, LAMPU INDIKATOR SINYAL MERAH TERUS., HARGA MAHAL, PELAYANAN CUEK., DISURUH MENUNGGU TERUS., KAPOK BELI INI..'.
          'Beli paket pembayaran VA dah 2 hari ga masuk2 dah komplain tapi belum ada kejelasan haduuh no recommend dah servicenya',
          'Ribet amat skrg harus login trus'.
         'Isi kuotanya pakai bca virtual account,ngga masuk2 kuotanya,di mohon nunggu ntah sampai kapan!orang perlu kuotanya skrg,ngga profesional bgt',
         'Ini apasih wifi sering tidak bisa terhubung, sering logout, mau login lagi susah, anehh',
         'susah daftatnya, sudah beli modem pas daftar disurubeli modem lagi. lebih baik daftar manual. harga juga tidak sesuai di promo. mahalllll promonya saja yg murah pembelian paket tetap MAHAL',
         'JANGAN JUAL PRODUK SAMPAH BOSS..!! ORBIT MAHAL, JARINGAN CUMA BAGUS DI BULAN PERTAMA, APLIKASI EROR, SERING FORCE CLOSE, GANTI NAMA PERANGKAT YANG TERHUBUNG BISA TAPI SETELAH KELUAR APLIKASI (LOGIN ULANG) SAAT CEK
  PERANGKAT TERHUBUNG LANGSUNG EROR APLIKASI, PULSA TERPOTONG KUOTA BELUM MASUK, AKUN TIDAK TERDAFTAR SAAT LOGIN (PADAHAL NOMER ITU SUDAH BENAR), PROTES DI PLAYSTORE CU<u>MA NGOMONG SILAHKAN HUBUNGI INSTAGRAM (TIDAK DIRESPON) I</u>
   'iklan tidak bisa d close parah',
         'Selalu gagal login. Password masuk ke aplikasi mv orbit selalu salah padahal sudah beberapa kali ganti. 🙎 \u200dd'.
         'Sering logout, dan jika login harus masukan kartu SIM di hp untuk mendapatkan kode otp. Sangat ribet. Apalagi ketika saya berada jauh dari modem orbitnya.',
          'Anying benar pihak telkomsel ini,,paket makin lama makin mahal,,sinyal orbit makin sulit,,anji*9',
          'Bigung mau beli paket data di aplikasi slalu muat ulang muat ulang gimana mau isi payah',
         'Gak jelas!',
         'Orbit oh orbit aplikasimu ga sesuai janjimu, alat sering error jaringan juga ahhhhh Mending pake my Republik atau sejenis daripada orbit',
          'Kenapa saya tidak bisa login ke aplikasi my orbit ?? Disitu di bilang terjadi kesalahan sistem .. mohon di balas',
          'Pada saat akan membuka perangkat yang terhubung aplikasi tiba2 keluar sendiri kecewa,harga paket naik asuu',
         'orbit makin mahal sekerang'.
          'Tolong deh di perbaiki lagi untuk costumer servicenya. Saya terkendala beli paket orbit sudah melakukan pembayaran tapi paket data belum bertambah. Sudah chat lewat si veronika gak ada respon. Uang 100ribu mungkin
  bukan uang banyak tapi seberapa banyak kalian ambil itu uang haram. Gak ikhlas saya seumur hidup'.
          'Sering long off dari hp yg terdaftar harus login ulang mengunakan kartu yg tertanam di orbit.',
          'Mau isi ulang kuota tapi error tolong diperbaiki di aplikasinya 🤚 '.
          'Semakin kesini semakun mahal dan lelet'.
          'Ya bener mi orbit pertama di pakai lancar lama2 paket nya mahal di tambah lagi jarang bisa di pakai sinyalny tmbh susah ngk bisa setabil nyesel q beli .',
         'Sangat tidak di sarankan untuk beli soalny cuma bylan pertama yg lancar akses internetnya dan setelah itu parah lemot masih lebih cepat kalau kaki seribu pakai sepatu',
         'Gak tau kenapa ya gak bisa login, padahal udah daftar nomornya. Aplikasi aneh',
         'Mau ganti imail nunggu satu abat link endak di kirim lamaaaaa banget respont nya'.
         'Keluar sendiri dari aplikasi selalu dan berulang kali, paketannya makin mahal ga kayak pas pertama skrng 200rb cuman dapat 130gb aja ga kayak dulu 150gb ga nyampe 200rb. Sinyal bagus pas pertama makin hari makin
  parah.'],
        dtype=object)
```

Structure vs Unstructure

score	thumbsUpCount	reviewCreatedVersion	at
5	0	0.69.2	2023-09-04 01:34:10
1	0	0.69.2	2023-09-03 23:54:14
1	4	0.67.16	2023-09-03 20:19:18
3	0	0.69.2	2023-09-03 13:35:54
1	0	0.69.2	2023-09-03 12:52:15
5	1	0.8.23	2020-05-25 18:34:40
5	2	0.8.20	2020-05-22 05:17:46
4	0	None	2020-05-19 21:47:11
3	3	0.8.20	2020-05-19 17:17:52
5	1	0.8.20	2020-05-18 22:18:02

Get Real Data

MyOrbit - Internet Rumah WiFi

Telekomunikasi Selular







Tambahkan ke wishlist

Aplikasi ini tersedia untuk perangkat Anda



Tentang aplikasi ini →

Aplikasi eksklusif untuk pelanggan Telkomsel Orbit, internet rumah serba digital. Tingkatkan produktifitas Anda di rumah dengan kecepatan internet yang stabil.

Modem Anda juga bisa sederhana.

Mulai dari kontrol jumlah perangkat yang terhubung, atur waktu internetan dan cek pemakaian per hari, isi ulang kuota dan masih banyak lainnya.

Semua dari satu aplikasi....

Diupdate pada



Dukungan apl ~

Aplikasi serupa →



J&T Express Indonesia PT. Global Jet Express



Mitra Bukalapak: Pulsa, PPOB PT Bukalapak.com



Lite Anti Blokir Proxy Browser OZZMO

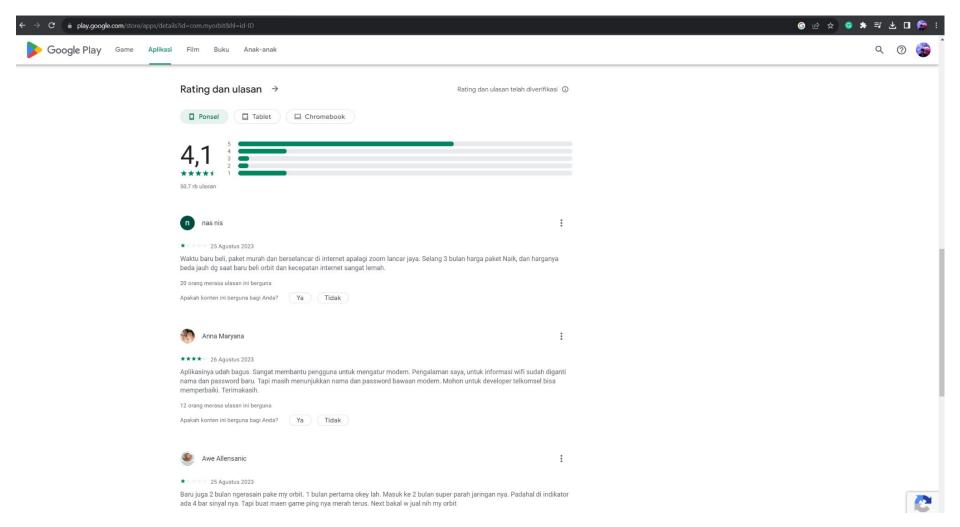


CGV CINEMAS INDONESIA CGV Cinemas 4.7★



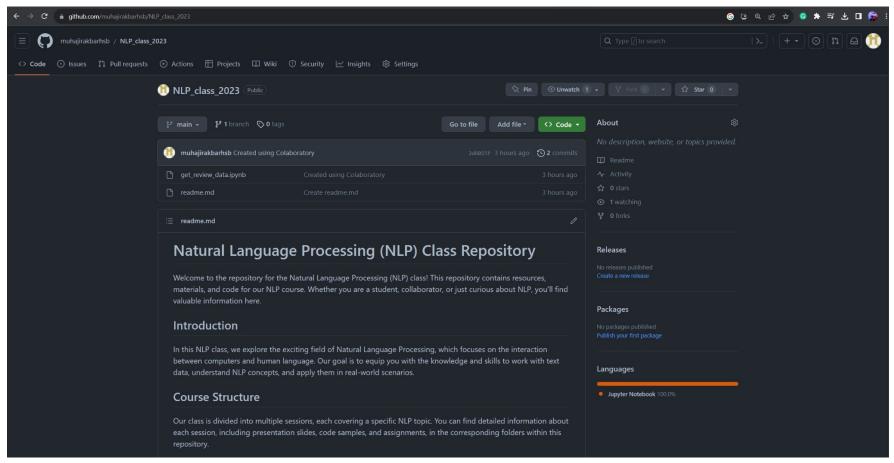
Shipper-Aplikasi Kirim Barang DT Objecteds Telescipal Lacistic





Python Time

https://github.com/muhajirakbarhsb/NLP_class_2023



Python Time

```
in python make a variable
word1 = "statistics"
word2 = "Statistics"
```

word1 == word2

what happened? anyone can conclude?

Class Meeting 1: Introduction to NLP (1 session) Introduction to NLP Basics of text data and its characteristics The importance of NLP in today's world Overview of the course structure and objectives	Class Meeting 7: Introduction to Topic Modeling(1 session) Introduction to topic Modelingg Topic Modeling Implementation LDA(Latent dirichlet allocation) Hands-on exercises with Topic Modeling	Class Meeting 14-15: Project Work (2 sessions) Dedicated sessions for students to work on NLP projects with guidance and assistance. Class Meeting 16: Project Presentations and Conclusion (1 session) Students present their NLP projects Recap of key takeaways from the course
Class Meeting 2: Text Preprocessing (1 session) Understanding the text preprocessing pipeline Tokenization, stemming, and lemmatization Stop words removal Hands-on exercises with Python for text preprocessing Class Meeting 3: Text Representation (1 session) Introduction to text representation techniques Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) n-gram Word embeddings (Word2Vec, GloVe)**just an introduction Practical exercises on text representation	Class Meeting 8: Mid-term Exam Class Meeting 9: Introduction to Word Embeddings (1 session) Fundamentals of word embeddings Word2Vec, GloVe, and FastText Word embedding applications Hands-on exercises with word embedding Class Meeting 10: Attention Mechanisms (1 session) Introduction to attention mechanisms Self-attention and multi-head attention Transformers architecture	Discuss further resources for NLP enthusiasts Course conclusion and feedback
Class Meeting 4: Introduction to ML, DL Supervised and Unsupervised (1 session) Introduction to deep learning (ML) Introduction to deep learning (DL) DL vs. traditional machine learning (ML) Machine Learning and their applications Neural networks and their applications Basic ML DL concepts and terminology	Practical examples of attention in NLP Class Meeting 11: Transformer Models (1 session) In-depth study of the Transformer model Pre-trained transformer models (BERT, GPT-2) Fine-tuning transformers for NLP tasks Transformer-based applications	
Class Meeting 5: Sentiment Analysis (1 session) What is sentiment analysis? Data collection and labeling for sentiment analysis Building a sentiment analysis model Practical sentiment analysis examples	Class Meeting 12: Advanced NLP Topics (1 session) Advanced NLP topics such as BERT, XLNet, and RoBERTa Transfer learning in NLP Ethics in NLP Recent developments and trends in NLP	
Class Meeting 6: Text Classification (1 session) Introduction to text classification Binary and multi-class classification Building a text classification model Real-world text classification examples	Class Meeting 13: Advanced NLP Topics and Deployment Process session) Introduction to LLM Introduction how industry utilize NLP to generate revenue Introduction MLOps for NLP(bonuses from practitioners)	in Industry (1

Terima Kasih