|  |  |
| --- | --- |
|  | **PROYEK AKHIR** |

**CHATBOT LAYANAN INFORMASI PENS**

**DENGAN REINFORCEMENT LEARNING**

**Muhammad Adryan Prawira Kusuma**

**NRP. 3120600010**

**DOSEN PEMBIMBING**

**Entin Martiana, S.Kom., M.Kom**

**NIP. 197403122000122001**

**Aliridho Barakbah, S.Kom., Ph.D.**

**NIP. 199009282022032009**

**PROGRAM STUDI SARJANA TERAPAN**

**TEKNIK INFORMATIKA**

**DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER**

**POLITEKNIK ELEKTRONIKA NEGERI SURABAYA**

**2024**

|  |  |
| --- | --- |
| logo pens kecil.png | **PROYEK AKHIR** |

PROGRAM STUDI SARJANA TERAPAN

TEKNIK INFORMATIKA

DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER

POLITEKNIK ELEKTRONIKA NEGERI SURABAYA

2023

**CHATBOT LAYANAN INFORMASI PENS**

**DENGAN REINFORCEMENT LEARNING**

**Muhammad Adryan Prawira Kusuma**

NRP. 3120600010

**DOSEN PEMBIMBING:**

**Entin Martiana, S.Kom., M.Kom.**

NIP. 197403122000122001

**Aliridho Barakbah, S.Kom., Ph.D.**

NIP. 197308162001121001

**PROGRAM STUDI SARJANA TERAPAN**

**TEKNIK INFORMATIKA**

**DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER**

**POLITEKNIK ELEKTRONIKA NEGERI SURABAYA**

**2024**

# KATA PENGANTAR

Segala puji bagi Allah yang telah memberikan nikmat dan menggerakan hati dan tangan ini dalam melakukan penelitian dengan judul “**Chatbot Layanan Informasi PENS dengan *Reinforcement Learning***”. Penelitian ini yang telah dilakukan selama satu tahun telah usai dengan ditandai dengan ditulisnya buku ini sebagai syarat maju sidang dan kelulusan dari program studi Sarjana Terapan Teknik Informatika.

Penelitian selama satu tahun meninggalkan kesan yang berarti yang turut mengembangkan ilmu pengetahuan dan pengalaman bagi diri ini dan bagi masyarakat umum. Tentunya selama perjalanan penelitian terdapat banyak rintangan dan halangan yang membuat diri ini tak dapat melewatinya kecuali dengan bantuan orang lain dalam menyelesaikannya. Hal ini membuktikan bahwa diri ini sungguh sangat miskin pengetahuan dan kemampuan, sehingga ucapan terima kasih saya haturkan pada :

* Allah *Subhanahu Wata’ala*  yang telah melimpahkan kemudahan kepada penulis sejak dalam pelaksanaan penelitan sampai dengan penyusunan buku Proyek Akhir
* Bapak M.Udin Harun Al Rasyid, S.Kom., M.Sc., Ph.D.
* Bapak Ahmad Syauqi Ahsan, S.Kom., M.T.
* Bu Desy Intan Permatasari, S.Kom., M.Kom
* Ibu Entin Martiana Kusumaningtyas S.Kom., M.Kom. selaku dosen pembimbing pertama yang telah meluangkan waktu dan memberi bimbingan, saran dan petuah dalam proses pengerjaan Proyek Akhir.
* Bapak Ali Ridho Barakbah, S.Kom., Ph.D. selaku dosen pembimbing kedua yang telah meluangkan waktu dan memberi bimbingan, saran dan petuah dalam proses pengerjaan penelitian Proyek Akhir.
* Teman teman kelas 4 D4 IT A, Keluarga Mahasiswa PENS, Pemandu Arunika, Himpunan Mahasiswa Teknik Informatika, UKM Tae Kwon Do PENS, Beswan Djarum 38, yang senantiasa menemani dan memberikan warna dalam periode perkuliahan ini mulai dari senang, sedih, susah bersama.
* Serta pihak-pihak lain yang telah membantu kami namun tidak dapat disebutkan satu persatu.

Demikian buku Proyek Akhir ini disusun. Kami menyadari masih banyak kekurangan dalam penyusunan buku ini. Kami berharap semoga apa yang

kami kerjakan dan kami tulis dalam laporan ini dapat membantu memberi bagi pihak-pihak terkait.

Surabaya, 18 April 2024

Muhammad Adryan Prawira Kusuma

# ABSTRAK

Untuk meningkatkan pelayanan Kemahasiswaan Bidang III kepada mahasiswa Politeknik Elektronika Negeri Surabaya membuat chatbot sebagai pusat layanan informasi mahasiswa untuk mendapatkan informasi seputar perkuliahan,keorganisasian,beasiswa,prestasi dan perlombaan. Penelitian ini bertujuan untuk membangun chatbot Layanan Informasi PENS berbasis *Reinforcement Learning (RL).*

*Reinforcement Learning*(RL) merupakan salah satu cabang kecerdasan buatan yang sedang berkembang dan banyak diteliti.Perbedaan dari RL dengan cabang lain pada studi kecerdasan buatan adalah algoritma pada RL tidak belajar dari korelasi antara masukan dan keluaran yang seharusnya. Namun algoritma ini belajar dari feedback atau timbal balik yang dia dapat saat dia memberikan suatu keluaran atau aksi dari masukan yang ia terima. Dengan metode pembelajaran ini, algoritma RL bisa menyelesaikan berbagai masalah yang penyelesaiannya tidak bisa diberikan langsung oleh manusia sehingga algoritma harus mencari sendiri dari permasalahanya, seperti kontrol pada robot, mengemudikan mobil, bermain game, hingga Chatbot.

Pada tugas akhir ini, penulis mengimplementasikan Reinforcement Learning yang dapat digunakan pada chatbot.Chatbot ini diaplikasikan pada platform Flask yang bisa berkomunikasi layaknya staff Kemahasiswaan. Chatbot ini dibangun dengan teknologi text mining untuk menampilkan informasi seputar Kemahasiswaan PENS Bidang III

**Keywords :** Kemahasiswaan Bidang III, Chatbot, Reinforcement Learning, Text Mining

# DAFTAR ISI

[KATA PENGANTAR I](#_Toc165299221)

[ABSTRAK III](#_Toc165299222)

[DAFTAR ISI V](#_Toc165299223)

[DAFTAR GAMBAR VIII](#_Toc165299224)

[DAFTAR TABEL IX](#_Toc165299225)

[BAB 1 1](#_Toc165299226)

[1.1 LATAR BELAKANG 1](#_Toc165299227)

[1.2 PERMASALAHAN 2](#_Toc165299228)

[1.3 TUJUAN 3](#_Toc165299229)

[1.4 MANFAAT 3](#_Toc165299230)

[1.5 METODOLOGI 3](#_Toc165299231)

[1. Studi Literatur 3](#_Toc165299232)

[2. Pengumpulan Data 4](#_Toc165299233)

[3. Perancangan Sistem 4](#_Toc165299234)

[4. Tahap Implementasi Sistem 4](#_Toc165299235)

[5. Tahap Pembuatan Laporan 4](#_Toc165299236)

[SISTEMATIKA PENULISAN 4](#_Toc165299237)

[BAB 2 6](#_Toc165299238)

[2.1 DESKRIPSI PERMASALAHAN 6](#_Toc165299239)

[2.2 TEORI PENUNJANG 6](#_Toc165299240)

[2.2.1 Machine Learning 6](#_Toc165299241)

[2.2.2 Reinforcement Learning 9](#_Toc165299242)

[2.2.3 Chatbot 10](#_Toc165299243)

[2.2.4 Text Mining 11](#_Toc165299244)

[2.2.5 Term Frequency Inverse Document Frequency 12](#_Toc165299245)

[2.2.6 Cosine Similiarity 13](#_Toc165299246)

[2.2.7 Open-Ai(GYM) 14](#_Toc165299247)

[2.2.8 Python 14](#_Toc165299248)

[2.2.9 Numpy 14](#_Toc165299249)

[2.2.10 Pandas 14](#_Toc165299250)

[2.3 PENELITIAN TERKAIT 15](#_Toc165299251)

[2.3.1 Pengembangan Aplikasi Whatsapp Chatbot untuk pelayanan akademik di Perguruan Tinggi 15](#_Toc165299252)

[2.3.2 Implementasi Natural Language Processing pada Chatbot Untuk Layanan Akademik 16](#_Toc165299253)

[2.3.3 Chatbot Gangguan Mental Berbasis Incremental Knowledge dengan Metode Temporal Different Learning 18](#_Toc165299254)

[2.3.4 Research Paper on Chatbot Development for Educational Institute 18](#_Toc165299255)

[BAB 3 22](#_Toc165299256)

[DESKRIPSI SISTEM 22](#_Toc165299257)

[3.1 DESKRIPSI SOLUSI 22](#_Toc165299258)

[3.2 RANCANGAN DESAIN SISTEM 22](#_Toc165299259)

[3.2.1 Information Resources 23](#_Toc165299260)

[3.2.3 Web Service 23](#_Toc165299261)

[3.2.2 Data Storage 24](#_Toc165299262)

[3.2.3 Text Processing 24](#_Toc165299263)

[3.2.3.1 Text Mining 25](#_Toc165299264)

[3.2.4 Reinforcement Learning 26](#_Toc165299265)

[3.3 RANCANGAN INFORMATION RESOURCES 27](#_Toc165299266)

[3.2.1 Dataset Layanan Informasi PENS 27](#_Toc165299267)

[3.2.5 Dataset Stopword List 28](#_Toc165299268)

[3.4 RANCANGAN DATABASE 29](#_Toc165299269)

[3.3.1 Tabel Reinforcement Learning 29](#_Toc165299270)

[3.3.2 Tabel Pertanyaan 30](#_Toc165299271)

[3.3.3 Tabel Jawaban 30](#_Toc165299272)

[3.4 USER INTERFACE 31](#_Toc165299273)

[3.2.1 Frontend System 32](#_Toc165299274)

[3.2.2 Backend 33](#_Toc165299275)

[BAB 4 34](#_Toc165299276)

[4.1 PARAMETER EKSPERIMEN 34](#_Toc165299277)

[4.2 KARAKTERISTIK DATA 34](#_Toc165299278)

[4.3 TEMPAT UJICOBA 34](#_Toc165299279)

[4.5 SPESIFIKASI PERALATAN UJICOBA 35](#_Toc165299280)

[4.5.1 Perangkat Keras 35](#_Toc165299281)

[4.5.2 Perangkat Lunak 35](#_Toc165299282)

[4.6 PENGUJIAN TEXT PROCESSING 35](#_Toc165299283)

[4.6.1 Hasil Pengujian Text Processing 36](#_Toc165299284)

[4.6.2 Eksperimen Text Mining 37](#_Toc165299285)

[4.6.3 Eksperimen Text Vectorizing (TF-IDF) 40](#_Toc165299286)

[4.7 Eksperimen Pencarian Jawaban 41](#_Toc165299287)

[BAB 5 43](#_Toc165299288)

[Daftar Pustaka 45](#_Toc165299289)

# DAFTAR GAMBAR

[Gambar 2.1 Diagram Reinforcement Learning 8](#_Toc165068307)

[Gambar 2.2 Flowchart Aplikasi Chatbot 9](#_Toc165068308)

[Gambar 2.3 Flowchart Chatbot 12](#_Toc165154979)

[Gambar 2.4 Output Chatbot 13](#_Toc165154980)

# DAFTAR TABEL

[Tabel 2.1 Hasil Penelitian Chatbot 1 11](#_Toc165154978)

[Tabel 3.1 Dataset Layanan Informasi PENS 1 21](#_Toc165154981)

[Tabel 3.2 Dataset Stopwords List 23](#_Toc165154982)

[Tabel 3.3 Reinforcement Learning 24](#_Toc165154983)

[Tabel 3.4 Skema Database Pertanyaan 25](#_Toc165154984)

[Tabel 3.5 Skema Database Jawaban 25](#_Toc165154985)

[Tabel 4.1 Spesifikasi Device 28](#_Toc165154986)

[Tabel 4.1 Detail Pengujian Text Processing 29](#_Toc165154987)

[Tabel 4.2 Tabel Data Pengujian Case Folding 30](#_Toc165154988)

[Tabel 4.3 Sampel Data Pengujian Remove Punctuation 30](#_Toc165154989)

[Tabel 4.4 Sampel Data Pengujian Stripping 31](#_Toc165154990)

[Tabel 4.4 Sampel Data Tokenizing 31](#_Toc165154991)

[Tabel 4.5 Sampel Data Filtering 32](#_Toc165154992)

[Tabel 4.6 Sampel Data Stemming 32](#_Toc165154993)

[Tabel 4.7 Summary Text Mining 33](#_Toc165154994)

[Tabel 4.8 Eksperimen Text Vectorize 33](#_Toc165154995)

# BAB 1

**PENDAHULUAN**

## 1.1 LATAR BELAKANG

Peningkatan teknologi informasi dan komunikasi telah mengubah lanskap pendidikan di berbagai institusi akademik. Saat ini mahasiswa menghadapi tuntutan yang semakin kompleks, seperti aksesbilitas yang lebih tinggi terhadap sumber daya, peningkatan interaksi dengan kemahasiswaan dan staf administrasi[17]

Politeknik Elektronika Negeri Surabaya merupakan salah satu perguruan tinggi vokasi negeri yang menerapkan sistem informasi akademik yang memudahkan pengguna dalam pengelolaan akademik. Pelayanan informasi akademik disajikan dalam halaman portal website. Jenis pelayanan melalui portal website ini umumnya searah sehingga belum dimanfaatkan secara efektif dalam komunikasi dua arah.

Salah satu inovasi yang menjanjikan adalah penggunaan chatbot dalam konteks pelayanan akademik [18] . Chatbot adalah sebuah aplikasi yang sudah ada sejak tahun 1965 sampai dengan sekarang dengan menggunakan konsep tanya jawab.Chatbot banyak digandrungi sejumlah industri di Indonesia mulai dari jasa keuangan, ritel, e-commerce, hingga pemerintahan. Sebanyak 71% pengguna memanfaatkan chatbot untuk customer service, kemudian 12% untuk asisten virtual (menurut CEO & Kata.ai Irzan Raditya)[2]. Aplikasi chatbot menyederhanakan proses pengaksesan informasi yang dilakukan dengan menggunakan *website*. Pengguna hanya perlu menginputkan pertanyaan kepada chatbot kemudian pengguna akan langsung mendapatkan respon jawaban yang diinginkan dengan cepat.

Aplikasi chatbot yang dikembangkan pada penelitian ini berbasis Machine Learning yang terlatih yang mampu menyelesaikan tugas yang ada. Terdapat tiga kelas utama dari teknik Machine Learning, yaitu *Supervised learning, Unsupervised Learning,*  dan yang terakhir adalah *Reinforcement Learning.*

*Reinforcement Learning*(RL) merupakan salah satu cabang dari Machine Learning. Algoritma pada RL berjalan dengan menerima masukan dari observasi sebuah *environtment* dan memberikan aksi yang sesuai. Metode ini tidak belajar dari apa keluaran yang seharusnya ia keluarkan, namun metode ini belajar dari *feedback* yang duberikan dari *environtmentnya.* Berbeda dengan supervised learning yang belajar dari data yang diberikan. Metode RL bertujuan untuk memaksimalkan reward atau feedback yang diberikan environtment. Apabila dianalogikan, pembelajaran pada RL bekerja seperti padahalnya anak kecil yang sedang belajar berjalan. Pada awalnya tentu anak kecil akan jatuh saat pertama kali belajar berjalan, dan akan menerima *feedback* berupa rasa sakit saat jatuh.Tentu anak kecil yang sedang belajar ini akan berusaha agar tidak mendapat rasa sakit tersebut. Sama halnya pada RL, agen akan berusaha memaksimalkan reward atau memperkecil *punishment* yang didapat.

Dalam hal ini, penggunaan Reinforcement Learning yang akan menjadi solusi pada masalah tuntutan peningkatan mutu pelayanan informasi akademik untuk lebih efektif.

Tantangan Indonesia di era digital ini adalah menghasilkan lulusan yang dapat beradaptasi dengan perkembangan zaman. untuk menghasilkan lulusan di era digital ini tentu perlu peningkatan mutu pendidikan terutama di pendidikan tinggi. peningkatan mutu ini dapat dilakukan dengan lebih modern dalam pemberian layanan informasi seputar pelayanan akademik. untuk memungkinkan menjawab tantangan ini, diperlukan penyelenggaraan Chatbot Layanan Informasi yang diharapkan menunjang mutu pelayanan akademik di Institusi Politeknik Elektronika Negeri Surabaya. Implementasi Chatbot ini akan membantu staf administrasi akademik Politeknik Elektronika Negeri Surabaya untuk memberikan layanan informasi.

Dengan menerapkan penggunaan chatbot untuk mencari informasi akademik kemahasiswaan bidang III, Politeknik Elektronika Negeri Surabaya, adalah satu langkah untuk menunjang mutu pendidikan di pendidikan tinggi.

## 1.2 PERMASALAHAN

Politeknik Elektronika Negeri Surabaya masih menggunakan sistem konvensional dalam pertukaran informasi secara satu arah, dimana kebutuhan tenaga staf administrasi kemahasiswaan untuk menjelaskan informasi yang berulang kepada mahasiswa yang membutuhkan informasi dari pengajuan proposal kegiatan Himpunan, Lomba dan Beasiswa, secara cepat dan efisien. Melihat lokasi dari kemahasiswaan III PENS masih kurang strategis, dan Staf kemahasiswaan yang terbatas waktu pelayanan dalam melingkupi jam kerja kantor, dan *website* Kemahasiswaan diperlukan update konten berkala mengenai informasi pendaftaran Beasiswa dan Perlombaan.

Mahasiswa Politeknik Elektronika Negeri Surabaya perlu usaha lebih yang kurang efektif untuk mencari informasi tersebut dikarenakan terkadang informasi yang disediakan masih perlu diupdate secara berkala.

Ada beberapa solusi yang sudah diimplementasikan seperti adanya website https://kemahasiswaan.pens.ac.id/ atau mendatangi pihak bagian kemahasiswaan, dan PENS BLAST secara teknis menggunakan fitur*Broadcast* ke E-Mail PENS dan Telegram PENS, tetapi untuk penerapannya masih kurang efektif dikarenakan mahasiswa PENS memiliki interval lebih banyak membuka sosial media pihak ketiga dan mengunjungi pihak administrasi kemahasiswaan.

Permasalahan diatas dapat dikerucutkan bagaimana proyek akhir ini mengimplementasikan Teori Reinforcement Learning ke dalam Chatbot Layanan Informasi PENS yang berinteraksi dengan *user*  dan bagaimana mengintegrasikan *Reinforcement Learning*  dengan database.

## 1.3 TUJUAN

Berdasarkan permasalahan yang telah dipaparkan diatas, tujuan dari aplikasi baru berbasis chatbot untuk memberikan sarana pelayanan informasi Perguruan Tinggi yang tidak terbatas ruang, waktu dengan interaksi dua arah. Chatbot ini dibangun dengan Reinforcement Learning yang bertujuan mahasiswa Politeknik Elektronika Negeri Surabaya dapat berinteraksi layaknya berkomunikasi dengan Pelayanan Akademik sehingga memudahkan Mahasiswa untuk mendapatkan informasi akademik yang diinginkan.

Berikut adalah cakupan dari percakapan Chatbot Layanan Informasi PENS :

1. Prosedur Proposal kegiatan mahasiswa

2. Prosedur pengajuan surat rekomendasi

3. Prosedur pengajuan beasiswa

4. Prosedur pengajuan prestasi mahasiswa

5. Prosedur pembayaran UKT/SPP

6. Informasi profil Organisasi Mahasiswa, UKM, & Komunitas

## 1.4 MANFAAT

Pada penelitian ini diharapkan dapat memberikan sebuah sarana yang mempermudah mahasiswa untuk mengakses informasi terkait dengan perguruan tinggi. Selain itu, penelitian ini diharapkan memudahkan bagi perguruan tinggi dalam memberikan pengalaman pengguna yang lebih efektif ketika mencari informasi tentang perguruan tinggi.

## 1.5 METODOLOGI

Metodologi yang digunakan pada proyek akhir ini meliputi :

### Studi Literatur

Studi literatur merupakan langkah awal dalam pengerjaan proyek akhir ini. Pada penelitian ini menggunakan chatbot sebagai solusi dari permasalahan tersebut dan penggunaan Reinforcement Learning. Reinforcement Learning adalah suatu metode deep learning yang digunakan untuk memprediksi jawaban yang tepat untuk chatbot. Dengan menggunakan metode temporal different learning, chatbot tersebut akan memposisikan dirinya masih belajar dengan prediksi jawaban yang sudah disediakan.

### Pengumpulan Data

Pengumpulan data dilakukan pengumpulan data-data yang diperlukan untuk mendukung pengerjaan proyek akhir. Data yang dibutuhkan adalah informasi administrasi akademik Politeknik Elektronika Negeri Surabaya. Data ini digunakan sebagai acuan dalam menentukan akurasi informasi terkait yang dibutuhkan oleh pengguna. Data yang dibutuhkan adalah :

* + - 1. Prosedur Proposal kegiatan mahasiswa
      2. Prosedur pengajuan surat rekomendasi
      3. Prosedur pengajuan beasiswa,subsidi lomba
      4. Prosedur pembayaran UKT & IKOMA
      5. Informasi Profil Organisasi Mahasiswa & UKM
      6. Informasi Lomba & Beasiswa

### Perancangan Sistem

Pembuatan desain sistem untuk aplikasi terkait dengan Frontend, Backend, Model Reinforcement Learning.

### Tahap Implementasi Sistem

Dalam tahap ini akan dibangun arsitektur dan desain sistem dari sistem berdasarkan rancangan arsitektur dan desain aplikasi yang sudah didefinisikan sebelumnya

### Tahap Pembuatan Laporan

Tahap terakhir setelah dilakukan implementasi sistem yaitu pembuatan laporan mengenai penelitian pembuatan aplikasi ini.

## SISTEMATIKA PENULISAN

**Bab 1 Pendahuluan**

Bab ini berisi tentang latar belakang, perumusuan masalah, tujuan,manfaat, serta sistematika pembahasan dari penelitian ini

**Bab 2 Kajian Pustaka**

Bab ini berisi tentang deskripsi permasalahan, teori penunjang serta penelitian yang terkait dengan penelitian ini.

**Bab 3 Desain Sistem**

Bab ini membahas tentang deskripsi solusi serta desain sistem meliputi rancangan dan alur proses yang akan dilakukan untuk menyelesaikan masalah beserta penjelasannya.

**Bab 4 Eksperimen dan Analisis**

Bab ini menyajikan hasil analisa dari percobaan yang dilakukan dalam pembuatan proyek akhir ini.

**Bab 5 Penutup**

Bab ini berisi kesimpulan dari uji coba perangkat yang sudah dibuat, saran untuk pengembangan dan perbaikan serta penyempurnaan terhadap aplikasi yang dibuat.

# 

# BAB 2

**KAJIAN PUSTAKA**

## 2.1 DESKRIPSI PERMASALAHAN

Tingginya Tuntutan permintaan informasi seputar akademik di Politeknik Elektronika Negeri Surabaya meliputi prosedur pengajuan proposal kegiatan mahasiswa, surat rekomendasi, pengajuan beasiswa, prosedur prestasi mahasiswa masih menggunakan sistem konvensional satu arah melalui website dan kurangnya dokmentasi informasi oleh staf administrasi kemahasiswaan serta publikasi yang kurang meluas menyebabkan mahasiswa PENS tertinggal informasi mengenai lomba,beasiswa dan sarana pengembangan untuk mahasiswa itu sendiri.

Saat ini informasi-informasi tersebut didapatkan melalui website pens.ac.id yang masih konvensial menggunakan satu arah, penggunaan blast information melalui aplikasi pihak ketiga seperti Gmail,Telegram dan Zimbra PENS dimana informasi yang dibagikan hanya yang diterima oleh staf administrasi mahasiswa. Opsi ketiga mahasiswa mendatangi kantor bagian kemahasiswaan untuk mendapatkan informasi secara interaksi dua arah yang masih terbatas waktu jam kerja dan ketersediaan tenaga kerja staf.

Akan lebih memudahkan jika semua informasi tersebut dilewatkan melalui satu pintu yang sama melalui sebuah chatbot yang mana mencerminkan permintaan informasi dari mahasiswa ke perguruan tinggi terkait.

## 2.2 TEORI PENUNJANG

### 2.2.1 Machine Learning

Menurut buku berjudul “Machine Learning” yang ditulis oleh (Dinata, et al.), *Machine Learning*  merupakan sebuah ilmu komputer yang dapat bekerja tanpa harus diprogram oleh manusia. Dilansir dari dicoding.com, pembelajaran mesin tersebut dikembangkan berdasarkan beberapa disiplin ilmu, yaitu statistika, matematika dan *data mining* sehingga melalui pembelajaran tersebut, mesin tidak perlu diprogram atau diperintah terlebih dahulu untuk melakukan sesuatu dengan menganalisa data. Penggunaan Machine Learning sudah banyak dikembangkan dan digunakan untuk otomasi proses dalam berbagai bidang. Menurut Jarot S Suroso Machine Learning terdapat banyak jenis dan tipe nya berdasarkan hasil yang akan didapatkan dengan melihat kondisi data yang diperoleh.

1. Supervised Learning

Supervised Learning (pembelajaran terarah) adalah salah satu metode pembelajaran mesin dimana hasil yang diharapkan pengguna, sudah diketahui atau dimiliki informasinya oleh sistem.

Hal ini berarti metode pembelajaran ini bekerja dengan memanfaatkan kembali data-data dan hasil output yang pernah dimasukkan oleh pengguna atau dikerjakan oleh sistem sebelumnya.

Pada metode ini, pola input dan pola output dibutuhkan untuk mengenali suatu informasi dalam bank memori, pola input dan pola output dibutuhkan untuk mengenali suatu informasi dalam bank memori. Ketika suatu pola input dibentuk, sistem akan meneruskan rangsangan data hingga ke bank memori dan sistem output. Sistem output yang menerima rangsangan data akan menampikan pola output dan mencocokan polanya dengan pola input. Apabila pola input dan pola output tidak ada yang cocok, maka output akan error. Dan jika nilai error cukup besar, pembelajaran lebih lanjut perlu dilakukan.

Beberapa contoh sistem algoritma yang menerapkan metode supervised learning adalah algoritma Hebbian(Hebb rule), algoritma perceptron, algoritma Adaline, algoritma Boltzman, algoritma Hapfield, dan algoritma Backpropagation.

Contoh studi kasus pemecahan masalah dengan metode Supervised Learning adalah misalnya kita ingin memilah email mana yang termasuk spam dan mana yang bukan. Saat pertama kali kita memutuskan suatu email dari pengirim tertentu adalah spam, sistem tidak memiliki data pemilahan sehingga semua email diterima sebagaimana mestinya. Namun setelah kita menandai email dari suatu pengirim adalah spam, sistem akan secara otomatis terus memasukan email tersebut ke folder spam sampai kita membatalkan stempel atau pilihan spam pada si pengirim email tersebut.

1. Unsupervised Learning

Unsupervised Learning ( pembelajaran tidak terarah) adalah metode dalam materi pembelajaran mesin. Konsep yang metode ini digunakan jauh berbeda dengan metode Supervised Learning dimana pada metode ini hasil yang diharapkan tidak dapat diketahui oleh siapapun. Dengan kata lain, hasil yang akan ditampilkan hanya bergantung kepada nilai bobot yang disusun pada awal pembangunan sistem dan tentu masih dalam ruang lingkup tertentu. Tujuan utama dari metode pembelajaran ini adalah agar para penggunanya dapat mengelompokkan objek-objek yang dinilai sejenis dalam ruang atau area tertentu. Metode pembelajaran ini sangat cocok digunakan untuk mencari atau mengklasifikasi suatu pola dari banyak objek sejenis yang tidak sepenuhnya sama.

Beberapa contoh sistem algoritma yang menggunakan Unsupervised Learning adalah algoritma kompetitif, algoritma Hebbian , algoritma Kohonen, algoritma Neocognitron.

Contoh studi kasus pemecahan masalah dengan metode Unsupervised Learning adalah misal suatu pusat perbelanjaan ingin melakukan bongkar terhadap satu truk berisi sepatu campur. Agar dapat dijual sepatu-sepatu tersebut perlu dikelompokkan brand dan ukurannya. Dalam hal ini, pihak pusat perbelanjaan tidak perlu memasukkan datanya terlebih dahulu karena data yang ada di lapangan saat itulah yang langsung diproses untuk mengelompokkan sepatu tersebut sesuai brand dan ukurannya.

1. Semi-Supervised Learning

Merupakan gabungan dari supervised learning dan unsupervised learning. Di sini dataset untuk pelatihan sebagian memiliki label dan sebagian tidak. Google Photos adalah contoh implementasi yang sering kita gunakan. Pada Google Photos kita bisa memberi tag atau label untuk setiap orang yang ada dalam sebuah foto. Alhasil, ketika kita mengunggah foto baru dengan wajah orang yang sebelumnnya sudah kita beri label, Google Photos akan secara otomatis mengenali orang tersebut.

Salah satu contoh dari model supervised learning adalah Deep Belief Network(DBNs). DBNs adalah model grafis dengan multiple layer yang dapat belajar teknik mengekstrak data training secara efisien. Dua jenis layer pada DBNs adalah visible atau input layer dan hidden layer.

Menurut Geron, DBNs berdasar pada komponen unsupervised yang disebut restricted Boltzmann machine(RBMs). RBMs dilatih secara berurutan dengan algoritma unsupervised learning, kemudian seluruh sistem disesuaikan dengan teknik supervised learning.

Campbell dalam tulisannya menyatakan bahwa pendekatan DBNs telah berhasil menyelesaikan pemodelan akustik pada speech recognition. DBNs menunjukkan sifat perkiraan yang kuat, peningkatan kinerja, dan merupakan parameter yang efisien.

1. Reinforcement Learning (yang akan digunakan oleh penulis)

Reinforcement Learning dikenal sebagai model yang belajar menggunakan sistem reward dan penalty. Menurut Winder, reinforcement learning adalah teknik yang mempelajari bagaimana membuat keputusan terbaik, secara berurutan, untuk memaksimalkan ukuran sukses kehidupan nyata. Entitas pembuat keputusan belajar melalui proses trial dan error

Reinforcement Learning memiliki empat komponen, yaitu action, agent, environtment, dan reward. Action adalah setiap keputusan yang diambil. Misal, saat kita berkendara, action yang kita lakukan adalah mengendalikan kemudi, menginjak gas, dan mengerem. Agent adalah entitas yang membuat keputusan, contohnya adalah perangkat lunak, atau robot, atau bahkan manusia. Environtment adalah sarana untuk berinteraksi, yang dapat menerima action dan memberikan respon berupa hasil maupun data berupa satu set observasi baru. Reward diberikan saat agent berhasil menyelesaikan tantangan. Mekanisme feedback ini membuat agent belajar tentang tindakan mana yang menyebabkan kesuksesan (menghasilkan reward), atau kegagalan (menghasilkan penalti). Keempat komponen ini merepresentasikan Markov decision proses(MDP).

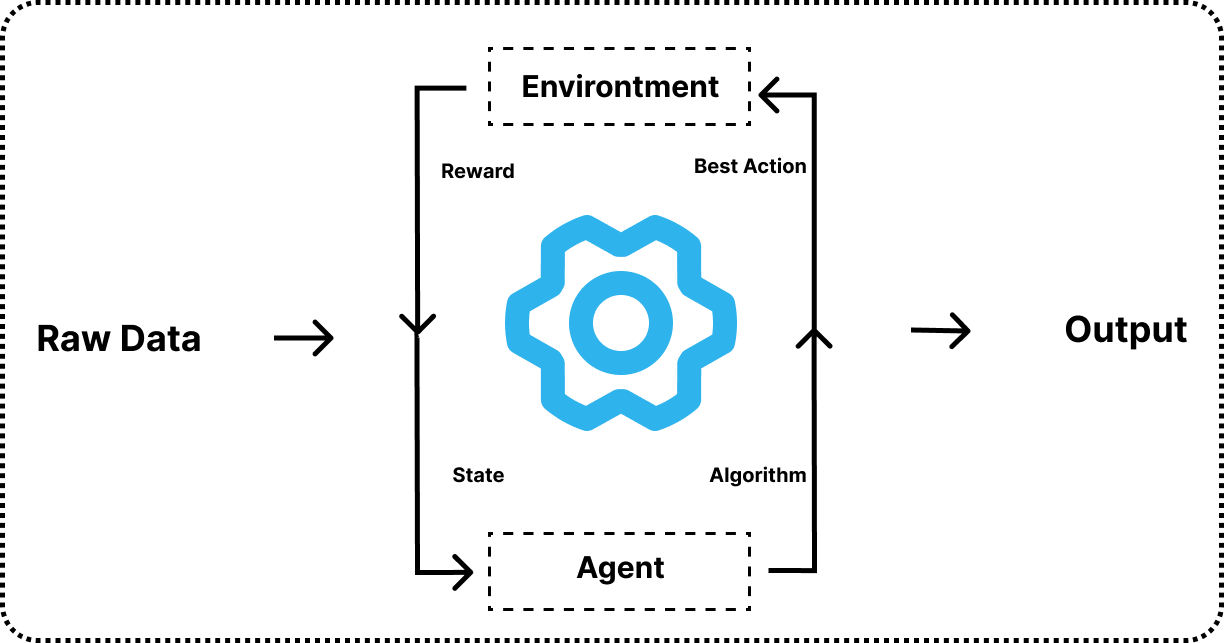
Model reinforcement learning belajar agar terus mendapatkan reward dan menghindari penalti. AlphaGo, sebuah program yang dikembangkan oleh Google DeepMind adalah contoh terkenal dari reinforcement learning. AlphaGo dibuat untuk mempelajari setiap langkah dalam jutaan permainan Go, untuk terus mendapatkan reward yaitu harus memenangkan permainan. AlphaGo terkenal setelah menjadi program komputer pertama yang berhasil mengalahkan seorang pemain Go Profesional yang juga merupakan juara dunia

### 2.2.2 Reinforcement Learning

Menurut Sutton dan Barto dalam bukunya “Reinforcement Learning: An Introduction”, Reinforcement Learning didefinisikan sebagai “pembelajaran melalui interaksi dengan lingkungan, di mana agen belajar untuk membuat keputusan yang optimal dengan memaksimalkan reward yang diberikan oleh lingkungan”

Reinforcement Learning adalah salah satu metode pembelajaran mesin yang memungkinkan mesin untuk belajar melalui interaksi dengan lingkungan. Tujuan utama dari Reinforcement Learning adalah untuk memaksimalkan reward atau hadiah yang diberikan oleh lingkungan. Dalam Reinforcement Learning, mesin belajar melalui trial and error, di mana mesin mencoba berbagai tindakan dan memperoleh feedback dari lingkungan dalam bentuk reward atau punishment.

Reinforcement Learning belajar melalui interaksi dari lingkungan dengan cara melakukan kegiatan yang berbeda dan mengalami banyak pengalaman kegagalan dan dikatakan sukes Ketika memaksimalkan rewards yang diberikan. Agen ini tidak diberitahu aksi mana yang akan dilakukan[6].



Gambar 2.1 Diagram Reinforcement Learning

Sumber : <https://techvidvan.com/tutorials/reinforcement-learning/>

(diakses pada tanggal 31 Desember 2023 jam 12.22 AM WIB)

### 2.2.3 Chatbot

Program chatbot pertama ditulis oleh Joseph Weizenbaum, profesor MIT pada tahun 1966. Pada waktu itu chatbot dibuat masih sangat sederhana. Meskipun perkembangan kecerdasan buatan ini sangat pesat dan canggih, namun chatbot tetap mempertahankan kedudukannya dalam dunia Artificial Intelligence (AI)[3]. Chatbot merupakan salah satu bentuk aplikasi Natural Language Processing (NLP) (Nila & Afrianto,2015).

Percakapan yang terjadi antara komputer dengan manusia merupakan bentuk respon dari program yang telah dideklarasikan pada database program.

Respon yang dihasilkan merupakan hasil pemindaian kata kunci pada input user dan menghasilkan respon balasan yang dianggap paling cocok, atau pola kata-kata yang dianggap paling mendekati (Ridwan, 2013).

Chatbot terdiri dari komponen bot program dan komponen brain file. Bot program merupakan program utama pada chatbot yang akan mengakses input dari pengguna, melakukan parsing dan kemudian membawanya ke brain file untuk kemudian respon. Bot program terdiri dari komponen scanner dan parser. Brain file merupakan otak dari chatbot itu sendiri yang menentukan bagaimana cara chatbot berpikir dan akan memberikan respon. Brain file biasanya berupa file plain text. Brain file berfungsi sebagaimana tabel informasi pada kompilator bahasa pemrograman tingkat tinggi. Di dalam brain file inilah disimpan semua kosakata, kepribadian dan pengetahuan (knowledge) dari chatbot. Semakin banyak pengetahahuan yang dimilki chatbot maka akan semakin besar ukuran file dari brain file tersebut.

### 2.2.4 Text Mining

Text mining adalah proses dimana informasi dan pola dapat ditemukan dan dieksplorasi dalam data yang tidak berstruktrur. Proses membersihkan dan mengonversi data teks menjadi format yang dapat digunakan disebut pemrosesan awal teks, dan ini harus dilakukan sebelum Anda dapat menggunakan salah satu dari banyak teknik penambangan teks[4].

Pemrosesan bahasa alami (NLP) adalah komponen kunci dari proses ini, dan untuk menyiapkan data dengan benar untuk analisis, NLP biasanya menggunakan metode termasuk identifikasi bahasa, tokenisasi, penandaan part-of-speech, chunking, dan parsing sintaks. Saat persiapan teks selesai, teknik penambangan teks dapat digunakan untuk mengekstrak wawasan dari data. Di antara metode penambangan teks yang tersebar luas ini adalah:

1. Case Folding

Case folding adalah salah satu bentuk text preprocessing yang paling sederhana dan efektif untuk mengubah semua huruf dalam dokumen menjadi huruf kecil. Sementara itu, karakter lain yang bukan termasuk huruf dan angka, seperti tanda baca dan spasi akan diabaikan.

1. Remove Punctuation

Remove punctutation adalah salah satu langkah penting dalam

Penting dalam text preprocessing yang bertujuan untuk membersi

hkan dari karakter-karakter yang tidak relevan dan mempermudah

proses analisis teks selanjutnya, Proses remove punctuation biasa

nya dilakukan dengan cara menghapus tanda baca seperti koma,

titik, tanda seru, tanda tanya, tanda kurung, tanda petik, dan

sebagainya.

1. Stripping

Stripping adalah proses penghapusan spasi putih berlebih dari tek. Spasi putih berlebih meliputi spasi ganda, tab, newline character, dan karakter whitespace laainnya yang tidak diperlukan. Proses ini bertujuan untuk membuat teks lebih konsisten dan memudahkan proses selanjutnya

1. Tokenizing

Tokenizing adalah proses pemisahan teks menjadi potongan-potongan yang disebut token untuk kemudian dianalisa. Kata, angka, simbol, tanda baca, dan entitas penting yang dapat dianggap sebagai token.

1. Filtering

Filtering adalah tahap mengambil kata-kata penting dari hasil token. Kata umum yang biasanya muncul dan tidak memiliki makna disebut dengan stopword, Misalnya penggunaan kata penghubung seperti dan, yang, setelah dan lainnya. Penghilangan stopword ini dapat mengurangi ukuran index dan waktu pemrosesan.

1. Stemming

Stemming adalah proses mengubah kata ke bentuk dasarnya. Misal kata “mendengarkan”, “dengarkan”,”dengar”. Selain itu juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk yang berbeda karena mendapatkan imbuhan yang berbeda pula.

### 2.2.5 Term Frequency Inverse Document Frequency

For a term  *i* in document j:

*Keterangan*

*t f = banyaknya term i pada sebuah dokumen*

*idf =Inverse document frequency*

*D = dokumen ke-d*

*df = banyak dokumen yang mengandung term i*

*t = term ke-t dari dokumen*

*W = bobot dokumen ke-d terhadap term ke-t*

Metode Term Frequency-Inverse Document Frequency (TF-IDF) adalah salah satu Teknik yang digunakan dalam pengolahan teks dan pemodelan Bahasa alami. Tujuan utama dari metode TF-IDF adalah untuk mengevaluasi seberapa penting suatu kata (term) dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar.

Konsep *Reinforcement Learning* dibagi beberapa elemen yaitu

*Input Raw Data, Agent,* dan *Environtment,* dimulai dari input raw data adalah berupa dataset dari pertanyaan yang diperoleh dari pertanyaan umum seputar Kemahasiswaan bidang III PENS yang akan diolah oleh Machine Learning agar dapat menjadi output model.

Data yang didapat akan diterima oleh *Agent. Agent*  memiliki peran sebagai pelaku utama menggunakan pendekatan algoritma yang digunakan saya kali ini adalah TF-IDF.

*Environtment*  berperan sebagai lingkungan untuk melatih agent dan memberi reward serta punishment, dengan analogi tindakan yang dilakukan benar sesuai dengan keakuratan teks chat yang diminta maka diberi reward hingga looping sesuai berapa kali diset dari awal training, apabila salah maka agent akan diberikan punishment oleh Environtment hingga seterusnya menghasilkan Model yang akurat akan keluar pada Output.

Output berbentuk model tersebut akan dideploy pada cloud sehingga dapat digunakan oleh user.

### 2.2.6 Cosine Similiarity

Cosine similiarity adalah sebuah metode yang digunakan untuk mengukur tingkat kesamaan dan kemiripan antara dua vektor dalam ruang multidimensi. Dalam chatbot, cosine similiarity digunakan untuk membandingkan kesamaan antara input pengguna dengan pola-pola yang ada dalam knowledge base atau database chatbot

Ketika diplot pada ruang multidimensi, dimana setiap dimensi berhubungan dengan kata dalam dokumen dan bukan besarnya. Karena unutk menghitung besarnya, yang harus dilakukan adalah menghitung jarak Euclidean. Cosine Similiarity sangat baik untuk mencari kemiripan antara dua dokumen berdasarkan kata yang muncul, karena meskipun dua dokumen yang mirip berjauhan berdasarkan jarak Euclidean karena ukurannya, dokumen-dokumen tersebut masih bisa memiliki sudut yang lebih kecil jika dibandingkan. Semakin kecil sudutnya, semakin tinggi nilai kemiripannya. Rumus untuk Cosine Similiarity adalah.

Keterangan :

A = Vektor A, yang akan dibandingkan kemiripannya

B = Vektor A, yang akan dibandingkan kemiripannya

A • B = dot product antara vektor A dan Vektor B

|A| = panjang vektor A

|B| = panjang vektor B

|A|B| = cross product antara |A| dan |B|

### 2.2.7 Open-Ai(GYM)

OpenAI Gym adalah toolkit open-resource yang dikembangkan oleh OpenAI untuk memfasilitasi pengembangan dan pengujian algoritma dalam konteks reinforcement learning. Gym menyediakan lingkungan simulasi yang beragam dan umum digunakan untuk membantu penelitian dan pengembangan dalam bidang machine learning dan reinforcement learning. Fungsi utama dari OpenAI Gym adalah menyediakan antarmuka konsisten dan masalah kebijakan. Dengan menggunakan Gym, pengembang atau peneliti dapat menguji dan membandingkan kinerja algoritma pembelajaran mesin mereka di berbagai tugas.

### 2.2.8 Python

Python adalah bahasa pemrograman yang populer, Python adalah bahasa pemrograman yang populer. Python sering dimanfaatkan dalam pengembangan web, perangkat lunak, penelitian, dan system scripting. Python dapat digunakan untuk menangani data besar dan melakukan operasi matematika yang kompleks. Python bekerja di berbagai platform seperti Windows, Mac, Linux, Raspberry Pi, dan lain-lain. Python dirancang untuk mudah dibaca, yaitu memiliki sintaks yang sederhana dan menggunakan bahasa inggris

### 2.2.9 Numpy

*Numpy* merupakan salah satu library dalam Python. Library ini dapat digunakan untuk banyak case dalam data science. Dengan adanya library ini, maka tidak diperlukan lagi baris kode yang panjang untuk menjalankan program machine learning. *Numpy* sendiri merupakan singkatan dari *Numerical Python.* Pada umumnya penggunaan library ini untuk menghitung operasi matematika pada array.

### 2.2.10 Pandas

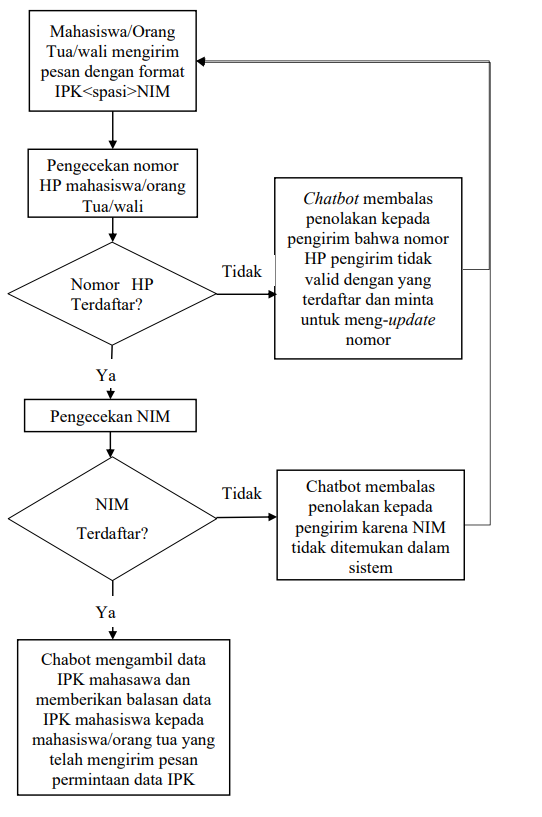
Pandas merupakan sebuah library pada *Python* yang berlisensi BSD dan open source yang menyediakan struktur data dan analisis yang mudah untuk digunakan. *Pandas* melakukan tugas penting seperti menyelaraskan data untuk perbandingan dan penggabungan dataset, penanganan data yang hilang, dan lain sebagainya.

Struktur data pandas dinamakan *datafram,* yaitu sebuah koleksi kolom berurutan dengan nama dan jenis. Dengan adanya fitur *dataframe*  memudahkan untuk membaca sebuah file dan menjadikannya sebuah table. *Dataframe* juga dapat mengolah suatu data dengan menggunakan operasi seperti *join,distinct, group by, agregasi,* dan fitur lainnya yang terdapat pada SQL. Format file yang didukung *Pandas*  meliputi file dengan ekstensi .txt, .csv, .tsv dan lainnya (Mutmainnah, 2019).

## 2.3 PENELITIAN TERKAIT

### 2.3.1 Pengembangan Aplikasi Whatsapp Chatbot untuk pelayanan akademik di Perguruan Tinggi

Penelitian ini dilakukan oleh Zulkifli Muhammad Hanif Mahasiswa Teknik Informatika Universitas Islam Indonesia.Pada penelitian ini ditampilkan untuk chatbot yang dapat membaca NIM,IPK,IPS serta no Hp mahasiwa yang bertujuan dapat meminta informasi berapa IPK/IPS dari mahasiswa terkait.



Gambar 2.2 Flowchart Aplikasi Chatbot

### 2.3.2 Implementasi Natural Language Processing pada Chatbot Untuk Layanan Akademik

Penelitian ini dilakukan oleh Alifya Hikmah dari Universitas Telkom Bandung. Penelitian ini menggunakan basis metode Natural Language Processing (NLP) yang terdiri dari case folding, tokenizing, 15 stemming obag of words. Setelah melalui preprocessing maka dilakukan proses training data menggunakan neural network unutk melatih dari hasil preprocessing.

Tabel 2.1 Hasil Penelitian Chatbot 1

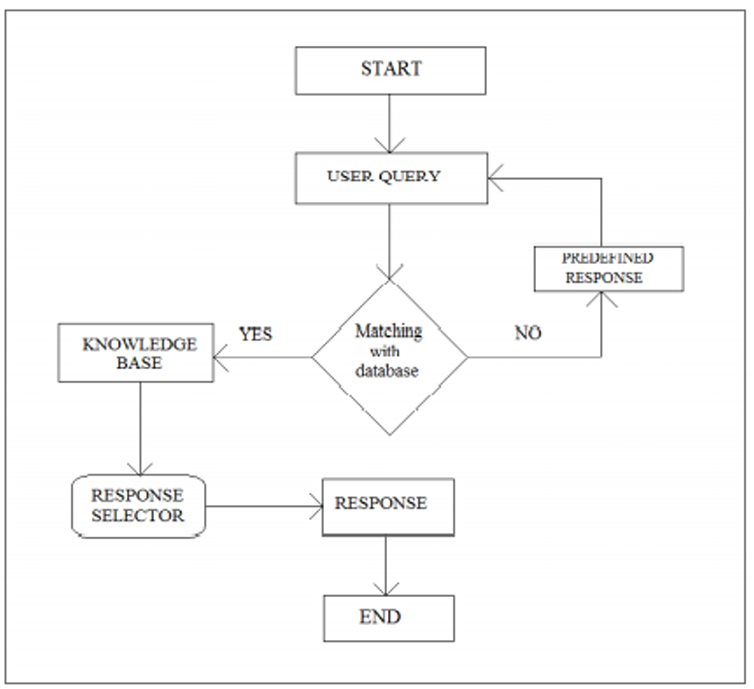
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Pertanyaan | Jawaban yang dihasilkan | Hasil yang diharapkan | Keterangan |
|  | Ada program beasiswa? | Kamu akses pada link berikut https://smb.telkomuniversity.ac.id |  |  |
|  | Download kalender akademik dimana | Informasi mengenai Kalender Akademik dapat kamu akses di link berikut, | Jawaban Sesuai | Valid |

### 2.3.3 Chatbot Gangguan Mental Berbasis Incremental Knowledge dengan Metode Temporal Different Learning

Penelitian ini dilakukan oleh Rifki Dwi Achsani Taqkwim[].Penelitian ini bertujuan untuk mengembangkan sebuah aplikasi edukasi kesehatan mental berbasis chatbot yang menggunakan metode Temporal Different Learning. Kesehatan mental merupakah masasah kompleks yang dipengaruhi banyak faktor, termasuk pengetahuan dan sikap masyarakat dalam meresponnya. Namun, masih banyak masyarakat yang kurang memahami kesehatan mental, malas mencari informasi di internet, dan kurang empati terhadap masalah tersebut. Untuk mengatasi tantangan ini, penelitian ini akan menggunakan aplikasi chatbot berbasis LINE Messengger yang menggunakan metode Temporal Different Learning. Chatbot ini akan dirancang agar dapat berkomunikasi layaknya dengan seorang dokter. Teknologi text mining akan digunakan unutk menampilkan informasi tentang penyakit mental berdasarkan masukkan pengguna. Aplikasi ini diharapkan dapat mempermudah masyarakat dalam mengedukasi diri mengenai kesehatan mental dengan respons yang efisien, tanpa terbatas oleh biaya, ruang, dan waktu.

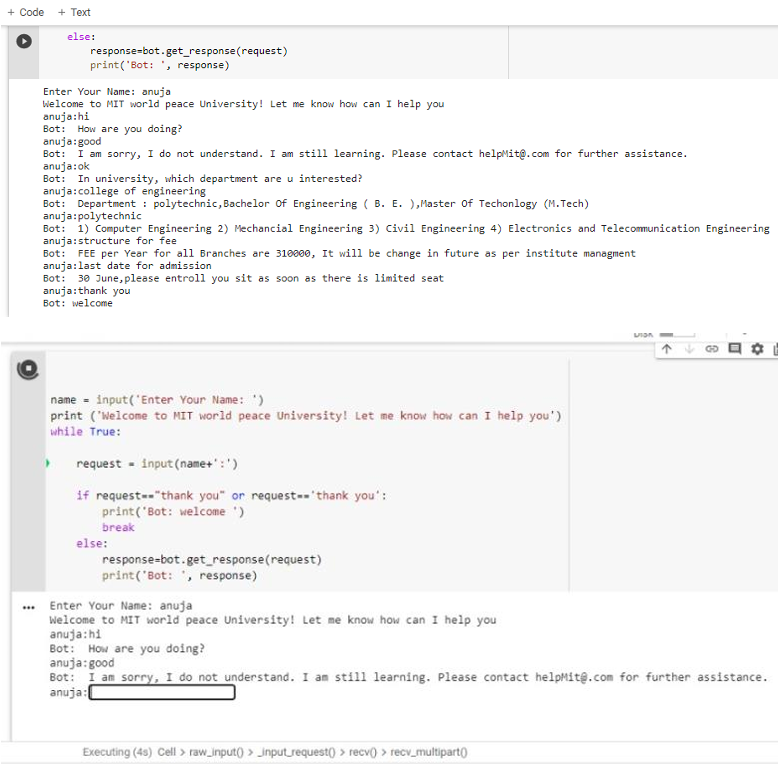
### 2.3.4 Research Paper on Chatbot Development for Educational Institute

Penelitian ini dilakukan oleh Dr. Vishwanath Karad dari MIT World Peace University,Pune,India.Penelitian ini membahas Chatbot sebagai kebutuhan untuk Institusi Pendidikan Proses system yang diajukan menggunakan library python yang mempermudah respon otomatis dari input pengguna.



Gambar 2.3 Flowchart Chatbot

Penelitian ini membuktikan keefektifannya. Dimana mengurangi manpower dan pengerjaan diatas kertas, dalam flowchart diatas juga dijelaskan mengenai alur dari aplikasi yang dibuat, dimulai dari user query yang diinputkan apakah sesuai dengan pertanyaan yang didatabase lalu akan diproses ke base knowledge dan diproses ke response selector, setelah response selector benar maka diberikan respon yang sesuai untuk diberikan kepada user.



Gambar 2.4 Output Chatbot

Output gambar diatas adalah percobaan dari eksesuki source code mengenai dari chatbot yang diciptakan oleh MIT World Peace University, user memberikan pertanyaan “Hi” , bot merespon “How are you doing?”, untuk testing dari output chatbot dilakukan pada platform Google Collab.

### 2.3.5 Pembangunan Aplikasi Chatbot Informasi Akademik berbasis Cosine Similiarity dan Library Sastrawi Stemmer (Studi Kasus: Teknik Informatika IT PLN).

Penelitian ini dilakukan oleh Rosida Nur Aziza, Tiara Sukma Ardanti, Efy Yosrita, Rahma Farah Ningrum[]. Penelitian ini bertujuan untuk menyediakan informasi akademik yang akurat dan cepat bagi mahasiswa Prodi Teknik Informatika IT PLN. Masalah yang dihadapi adalah ketidakakuratan dan ketidakjelasan informasi yang diterima melalui media sosial prodi atau dari sesama mahasiswa, serta keterlambatan dalam menjawab pertanyaan melalui chat. Untuk mengatasi masalah tersebut, penelitian ini mengembangkan sebuah aplikasi chatbot yang dapat memberikan jawaban otomatis terhadap pertanyaan yang masuk. Aplikasi chatbot ini beroperasi pada platform chat Telegram dan menggunakan library Sastrawi Stemmer untuk prapemrosesan pesan, metode TF-IDF untuk pembobotan kata, dan metode Cosine Similiarity untuk menghitung kemiripan antara objek. Penelitian ini melakukan pengujian untuk mengukur kinerja aplikasi chatbot. Rangkuman ini tidak menyediakan informasi tentang hasil pengujian yang telah dilakukan karena informasi tersebut tidak disediakan dalam abstrak yang diberikan.

# BAB 3

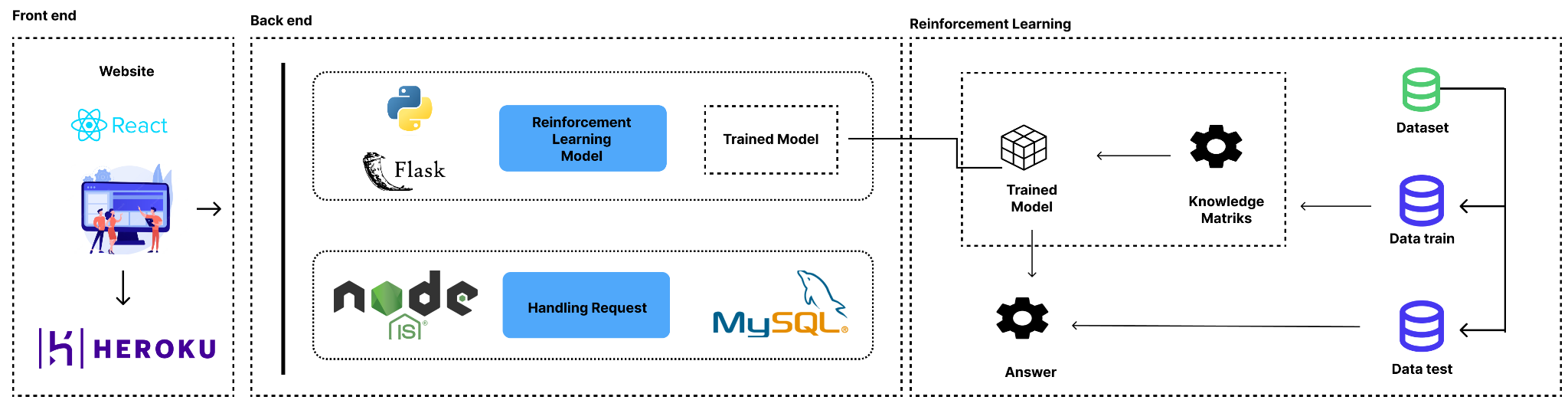
# DESKRIPSI SISTEM

## 3.1 DESKRIPSI SOLUSI

Proyek akhir ini dilakukan untuk mengembangkan sebuah aplikasi chatbot dengan metode Reinforcement Learning yang bertujuan untuk mempermudah dan mempercepat respon dari permintaan informasi kemahasiswaan PENS yang bergerak di Kemahasiswaan PENS Bidang III .

## 3.2 RANCANGAN DESAIN SISTEM

Berikut ini merupakan rancangan dari desain kerja system secara umum :

****

Gambar 3. 1 Diagram Sistem Chatbot Informasi Mahasiswa

Gambar diatas merupakan rancangan desain sistem dari aplikasi yang akan dikembangkan. Terdapat tiga sistem yang berkorelasi membangun aplikasi ini. Sistem tersebut diantaranya Frontend System, Backend System, dan Reinforcement Learning System. Untuk menghubungkan Frontend System dan Backend System akan menggunakan Restful API dimana Frontend System menjadi REST Client dan Backend System menjadi REST Server.

### 3.2.1 Information Resources

Tahap pertama penelitian ini dimulai dari information resources.

Pada tahap ini, dilakukan pengumpulan data yang sesuai akan disimpan ke dalam database yang nantinya dapat menentukan hasil dari aplikasi yang dikerjakan.

* **Wawancara pada Staff kemahasiswaan Politeknik Elektronika Negeri Surabaya**

Kegiatan ini bertujuan mengumpulkan informasi mengenai *SOP* administrasi akademik, informasi perlombaan, beasiswa, profil kampus, organisasi mahasiswa dan UKM.

* **Observasi Website** <https://kemahasiswaan.pens.ac.id> dan <https://pens.ac.id>

Website ini berisi kumpulan informasi mengenai SOP administrasi akademik, informasi perlombaan, beasiswa, profil kampus, organisasi mahasiswa dan UKM.

* **Observasi ZIMBRA PENS, PENS BLAST FORWARD SYSTEM**

Sistem ini berfungsi mengirimkan pesan Informasi lomba, pertukaran pelajar, beasiswa, webinar, secara masif melalui aplikasi pihak ketiga yaitu Zimbra,Gmail dan Telegram PENS.

* **Observasi Instagram Badan Eksekutif Mahasiswa PENS dan Himpunan Mahasiswa Teknik Informatika PENS**

Sosial media ini berisi kumpulan informasi mengenai lomba, pendampingan lomba. Subsidi lomba oleh hima, beasiswa, webinar, bootcamp hingga lowongan kerja.

### 3.2.3 Web Service

Pada Penelitian ini akan dibangun juga web service untuk menyimpan dan mendeploy aplikasi API. Untuk aplikasi web menggunakan Heroku

Pada dasarnya aplikasi ini memerlukan server yang digunakan untuk proses engine. Web service digunakan supaya untuk mengaksesnya tidak terpengaruh oleh platform. Ia akan menyediakan method-method yang dapat diakses oleh network. Jadi, Web Service dapat menghubungkan software yang memiliki platform berbeda. Pada proyek akhir ini menggunakan Flask dan Python sebagai web service.

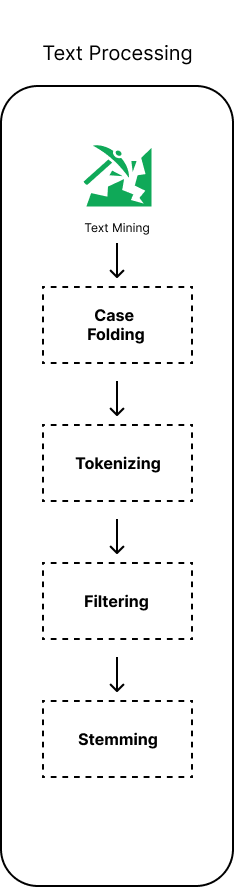
### 3.2.2 Data Storage

Data storage digunakan untuk menyimpan sumber data yang dimasukkan melalui Web Admin. Rancangan tersebut nantinya akan digunakan sebagai penyimpanan sumber data dan mempermudah untuk mencari dari sekumpulan sumber data. Untuk aplikasi chatbot ini menggunakan MySQL

### 3.2.3 Text Processing

Text Processing adalah proses yang dilakukan untuk mengolah data dari database dan input user yang nantinya akan digunakan untuk menentukan output yang diberikan kepada pengguna kedalam chatbot berupa informasi layanan akademik Politeknik Elektronika Negeri Surabaya

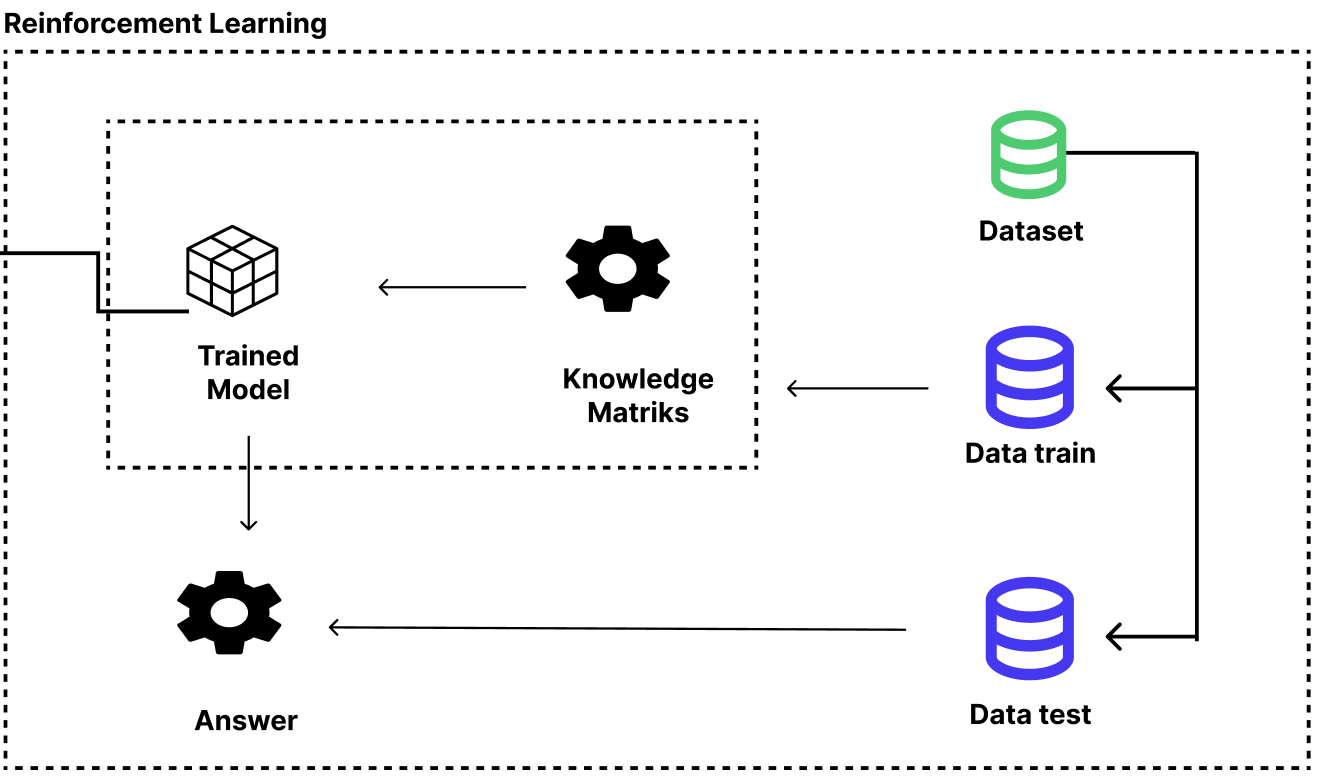
#### 3.2.3.1 Text Mining



*Text Mining*  adalah proses untuk mengolah data yang berasal dari database dan masukan user. Proses Text Mining telah dijelaskan pada poin 2.2.2. Text Mining terdiri dari proses tokenizing, filtering, dan stemming.

|  |  |  |
| --- | --- | --- |
| Nomor | Input User | Hasil Text Mining |
| 1 | Kegiatan UKM apa saja yang ada di PENS? | giat ukm pens |
| 2 | Beasiswa apa saja yang bekerja sama dengan PENS? | beasiswa pens |

### 3.2.4 Reinforcement Learning

****

Pada gambar 3.6, tabel Term Frequency (TF) yang terdiri dari keyword beserta frekuensi dari masing masing knowledge dipetakkan pada matriks besar dengan keyword sebagai kolom, id knowledge sebagai baris dan frekuensi sebagai nilainya. Apabila knowledge tidak memiliki keyword dari daftar keyword yang ada, maka nilai/frekuensinya diberi nilai default nol. Pada tabel 3.2 ditunjukkan contoh data Reinforcement Learning.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Kata-1 | Kata-2 | Kata-3 | Kata-4 | Kata-5 | Kata-n |
| Knowled  ge-1 | 5 |  |  |  |  |  |
| Knowled  ge-2 |  |  |  |  |  |  |
| Knowled  ge-3 |  |  |  |  |  |  |
| Knowled  ge-n |  |  |  |  |  |  |

Setelah membuat matriks besar seperti pada Tabel, akan ada opsi dimana pengguna bisa memilih ‘Ya’ atau ‘Tidak’, karena chatbot ini memposisikan masih belajar sesuatu. Ketika pengguna memilih jawaban ‘Ya’ maka nilai fitur yang diarahkan ke Knowledge tertentu akan diupdate dengan menggunakan metode Reinforcement Learning, berlaku juga proses jawaban ‘Tidak’.

## 3.3 RANCANGAN INFORMATION RESOURCES

### 3.2.1 Dataset Layanan Informasi PENS

Data ini berisi informasi seputar kumpulan Layanan informasi kemahasiswaan PENS.

Tabel 3.1 Dataset Layanan Informasi PENS

|  |  |
| --- | --- |
| **Informasi** | **Detail Informasi** |
| apa itu Politeknik Elektronika Negeri Surabaya? | Politeknik Elektronika Negeri Surabaya atau yang biasa dikenal dengan PENS merupakan salah satu perguruan tinggi negeri terbaik yang menyelenggarakan pendidikan vokasi pada bidang elektronik dan industri. Perguruan tinggi yang berkedudukan di Surabaya ini resmi berdiri pada tahun 1988. |
| Apa itu Gemastik? | GEMASTIK atau Pagelaran Mahasiswa Nasional Bidang Teknologi Informasi dan Komunikasi, merupakan program Pusat Prestasi Nasional, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi |
| Apa itu KMIPN? | Kompetisi Mahasiswa Informatika Politeknik Nasional KMIPN merupakan ajang bergengsi untuk Politeknik se-Indonesia di bidang Informatika. Pada tahun ini KMIPN akan dilaksanakan di Jakarta. |
| Apa saja Perlombaan yang diadakan tahun ini? | Terdapat GEMASTIK,KMIPN,PORSENI,PMW,P2MW |
| PENS berlokasi dimana? | Jalan Raya ITS - Kampus PENS Sukolilo Surabaya 60111, Indonesia |
| Kapan KMIPN dilaksanakan? | Tahap 1, Pendaftaran Perguruan Tinggi : 12 Mei - 10 Juli 2023    khusus Divisi II Keamanan Siber : 12 Mei 7 Juli 2023   Tahap 2, pendaftaran tim untuk semua divisi lomba : 12 Mei - 10 Juli 2023   khusus Divisi II Keamanan Siber : 12 Mei - 7 Juli 2023 |
| Kapan Periode Gemastik dilaksanakan? | Tahap 1, Pendaftaran Perguruan Tinggi : 12 Mei - 10 Juli 2023    khusus Divisi II Keamanan Siber : 12 Mei 7 Juli 2023   Tahap 2, pendaftaran tim untuk semua divisi lomba : 12 Mei - 10 Juli 2023    khusus Divisi II Keamanan Siber : 12 Mei - 7 Juli 2023 |

Tabel 3,1 merupakan contoh dataset yang dikumpulkan dari Observasi dan wawancara dengan Staf administrasi Politeknik Elektronika Negeri Surabaya dan website,tools yang disediakan untuk persebaran informasi.

### 3.2.5 Dataset Stopword List

Stopword List adalah kumpulan dari kata – kata yang tidak digunakan dalam sebuah kalimat. Data Stopword List akan digunakan untuk diproses pada tahap filtering pada proses Text Mining 2.2 yaitu menghilangkan kata – kata jika dalam sebuah kalimat terdapat kata pad Stopwords List.

Tabel 3.2 Dataset Stopwords List

|  |  |
| --- | --- |
| **No** | **Kata** |
| 1. | Dan |
| 2. | Dengan |
| 3. | Serta |
| 4. | Atau |
| 5. | Tetapi |
| 6. | Sedangkan |
| 7. | Namun |
| 8. | Sebaliknya |

Tabel 3.2 berisi sekumpulan data stopword list yaitu kata yang tidak dibutuhkan didalam sistem. Stopword list tentunya disesuaikan dengan kebutuhan sistem. Semua kata yang terdapat didalam stopword list akan dihapuskan didalam suatu kalimat yang akan diproses oleh sistem. Sehingga, akan mendapatkan hasil kata – kata yang penting saja yang akan diproses pada sistem.

## 3.4 RANCANGAN DATABASE

Database adalah kumpulan data, umumnya mendeskripsikan aktivitas satu organisasi yang berhubungan atau lebih. Database merupakan aspek yang sangat penting dalam sistem informasi dimana database merupakan gudang penyimpanan data. Selain digunakan sebagai penyimpanan data, database juga dapat mempermudah untuk mencari suatu informasi dari sekumpulan data yang tersusun dalam skala besar. Penerapan database biasanya dilakukan setelah melakukan analitis terhadap sistem yang ada.

### 3.3.1 Tabel Reinforcement Learning

Tabel Reinforcement Learning digunakan untuk menyimpan knowledge berupa macam – macam Informasi Layanan PENS beserta fitur – fitur yang digunakan untuk mencari istilah yang dituju. Berikut ini adalah gambar struktur tabel Reinforcement Learning yang ditunjukan pada tabel 3.3

Tabel 3.3 Reinforcement Learning

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Nama Kolom** | **Tipe data** | **Keterangan** |
| 1 | Level\_0 | Int8(64) | *Primary Key* |
| 2 | Reinforcement Learning | text | Menyimpan jawaban untuk chatbot |
| 3 | Fitur-n | Double precision | Menyimpan hasil nilai dari proses text mining, dan dijadikan sebuah fitur untuk chatbot |

### 3.3.2 Tabel Pertanyaan

Tabel pertanyaan digunakan untuk menyimpan knowledge dari history pertanyaan yang diajukan oleh user, bertujuan mencocokan dengan jawaban oleh respon chatbot dan diberikan skor reward dan punishment. Berikut ini adalah struktur tabel Pertanyaan yang ditunjukan pada tabel 3.6

Tabel 3.4 Skema Database Pertanyaan

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Nama Kolom** | **Tipe Data** | **Keterangan** |
| 1 | Id | Int8(64) | *Primary Key* |
| 2 | Pertanyaan | String | Menyimpan Pertanyaan untuk chatbot |

### 3.3.3 Tabel Jawaban

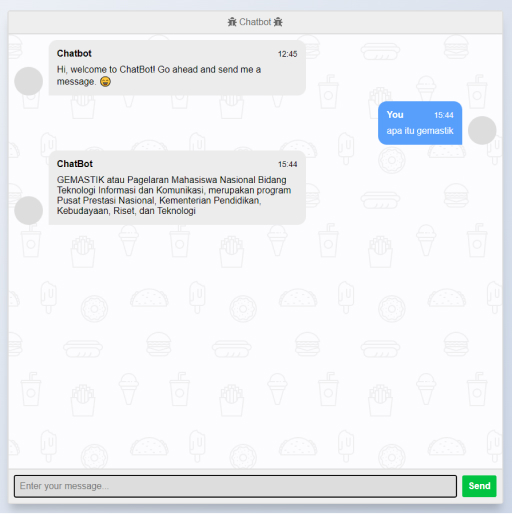
Tabel Jawaban digunakan untuk menyimpan knowledge dari respon jawaban yang diberikan, serta menampung skor berupa reward dan punishment dari User.

Tabel 3.5 Skema Database Jawaban

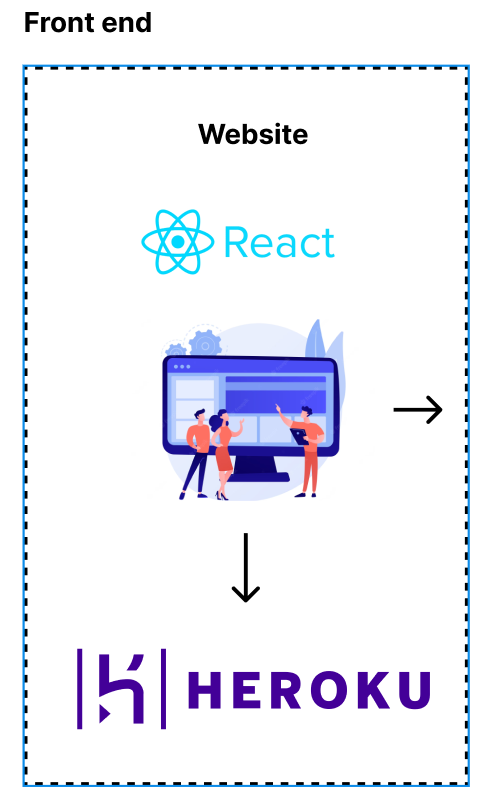
|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Nama Kolom** | **Tipe Data** | **Keterangan** |
| 1 | Id\_pertanyaan | Int8(64) | *Primary Key* |
| 2 | Jawaban | String | Menyimpan Jawaban untuk chatbot |
| 3 | Skor (Reinforcement Learning) | Float | Menampung Skor dari Reinforcement learning |

## 3.4 USER INTERFACE

Pada Proyek akhir ini terdapat perancangan user interface yang terdiri dari Website. Desain interface diharapkan dapat membuat pengguna nyaman dalam menggunakan aplikasi dan dapat meningkatkan pengguna dalam menggunakan aplikasi. Berikut adalah tampilan dari chatbot di Gambar.



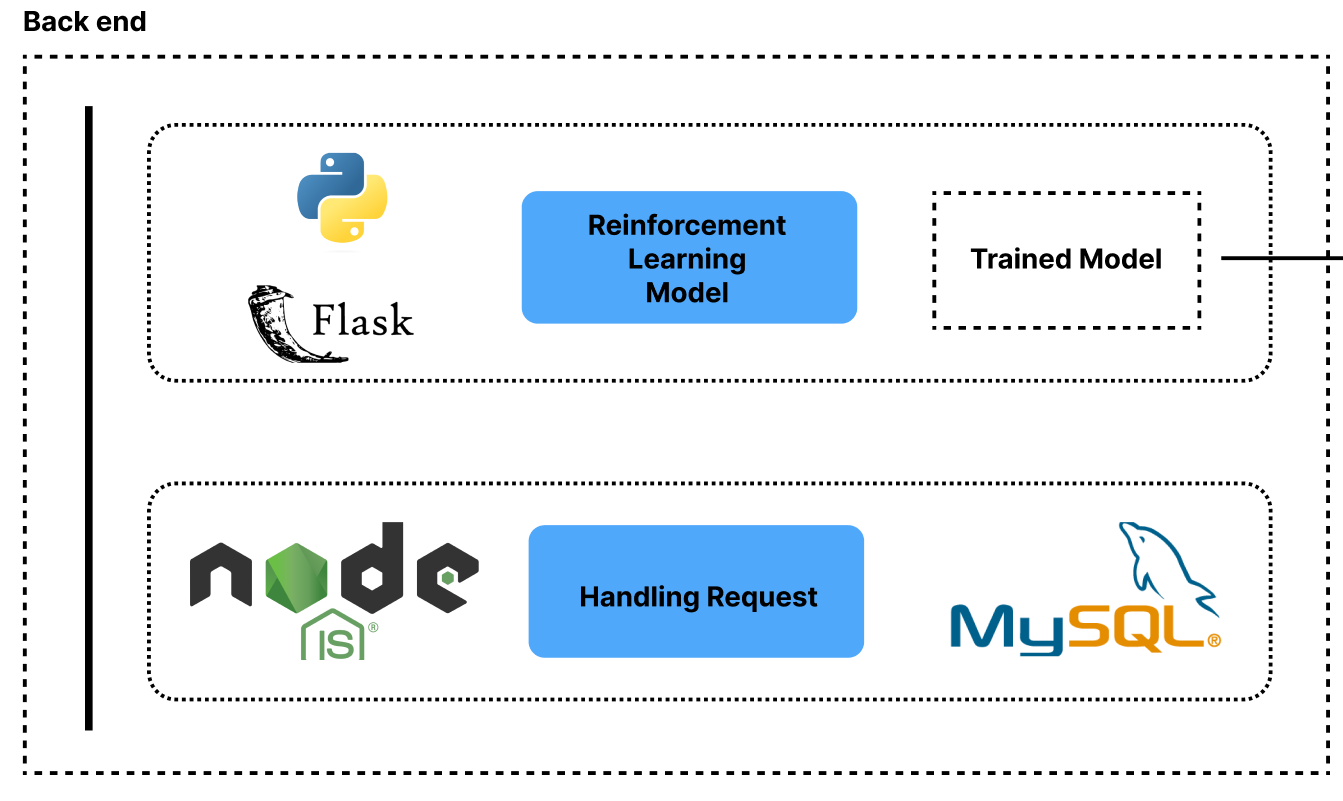
### 3.2.1 Frontend System



**Gambar 3.2 Frontend**

Sistem ini bertanggung jawab atas antarmuka dari clienside atau sisi pengguna, Frontend System yang digunakan pada aplikasi akan berjalan pada website. Pada sistem ini, pengguna dapat berinteraksi dengan chatbot mulai dari mengetikan pertanyaan hingga mendapat jawaban yang sesuai dengan keinginan pengguna seputar kemahasiswaan bidang III PENS.

### 3.2.2 Backend

****

**Gambar 3. 3 Backend**

Pada sistem ini bertanggung jawab untuk menyediakan Application Programming Interface (API) berupa Server untuk memproses response dan request. Request yang dimaksud adalah permintaan data yang ada pada server untuk dikembalikan lagi kepada client, yang biasanya disebut response. Backend System ini terbagi menjadi beberapa bagian yaitu API Gateway, Model Reinforcement Learning dan Question User Record.

# BAB 4

**EKSPERIMEN DAN ANALISIS**

## 4.1 PARAMETER EKSPERIMEN

Untuk mengukur tingkat keberhasilan sistem yang dibangun maka terdapat beberapa parameter eksperimen yang akan diujicobakan. Eksperimen dapat dikatakan berhasil jika dapat melewati beberapa tahap dengan baik. Tahap pertama adalah uji coba proses text preprocessing diantaranya adalah Tokenizing, Stemming, Filtering.

Selanjutnya akan dilakukan uji coba untuk proses mengubah teks dari string menjadi vector dengan menggunakan metode TF –IDF

Tahap yang terakhir yaitu mencari jawaban, pada penelitian ini akan dilakukan uji coba sistem dalam mencari jawaban dengan perhitungan cosine similiarity. Sistem dikatakan berhasil jika keyword dengan nilai tertinggi cosine similiarity yang dihasilkan memiliki jawaban yang sesuai dengan pertanyaan user.

## 4.2 KARAKTERISTIK DATA

Data yang digunakan sebagai masukan dalam melakukan eksperimen adalah data input dari user berupa pertanyaan berbentuk teks yang diinputkan melalui platform web. Data keluaran atau hasil dari eksperimen merupakan dari setiap proses baik text preprocessing ataupun perhtitungan jawaban. Data hasil uji coba text preprocessing merupakan sekumpulan keyword, sedangkan data hasil uji coba pencarian jawaban merupakan data response yang diambil dari database berupa teks dan dikirimkan kepada user melalui platform web

## 4.3 TEMPAT UJICOBA

Uji coba dilaksanakan di Laboratorium Knowledge Engineering PENS.

**4.4 WAKTU UJICOBA**

Uji coba dilaksanakan mulai dari bulan September 2023 sampai dengan Desember 2023

## 4.5 SPESIFIKASI PERALATAN UJICOBA

### 4.5.1 Perangkat Keras

Proses pembuatan, implementasi, dan kompilasi aplikasi dilakukan pada PC (Personal Computer). Spesifikasi PC yang digunakan ditunjukkan pada tabel berikut.

Tabel 4.1 Spesifikasi Device

|  |  |
| --- | --- |
| Aspek | Spesifikasi |
| CPU | AMD Ryzen™ 9 Mobile Processors with Radeon™ Graphics |
| RAM | 16.0 Gb DDR4 Dual Channel |
| GPU | Nvidia GeForce RTX 2060 6Gb |

### 4.5.2 Perangkat Lunak

Kemudian kebutuhan perangkat lunak serta library yang digunakan dalam pembuatan aplikasi adalah sebagai berikut :

* Python 3.12.0
* NLTK : Tokenize, corpus, stem
* Sastrawi : Stemmer
* Flask : 3.0.3
* Pandas : 2.2.1

## 4.6 PENGUJIAN TEXT PROCESSING

Pengujian Text Processing adalah pengujian yang dilakukan untuk mengetahui apakah input pengguna yang diolah dengan menggunakan metode text mining sudah mendapatkan output yang sesuai. Dari hasil pengujian Text Processing tersebut terdapat beberapa tahap mulai dari pengujian filtering (melakukan penyaringan kata- kata yang tidak dibutuhkan), stemming (mengubah kata – kata mnejadi kata dasar), dan pengecekan knowledge melalui Reinforcement Learning.

### 4.6.1 Hasil Pengujian Text Processing

Pengujian Text Processing merupakan pengujian pada dataset Layanan Informasi PENS, terutama pada data teks sehingga menjadi fitur yang dapat diproses pada tahapan selanjutnya.

Tabel 4.1 Detail Pengujian Text Processing

|  |  |
| --- | --- |
| Nama | **Data Processing** |
| Tujuan | Mengetahui perubahan data sebelum melalui data processing dan setelahnya. |
| Peralatan | * Laptop ASUS TUF FX506IV RYZEN 9 4900H |
| Waktu Pengujian | Agustus 2023 – Oktober 2023 |
| Parameter  Pengujian | * Ketepatan |

Pengujian pada tahap Data Processing terdiri dari beberapa tahap, seperti yang ditunjukkan pada skema berikut.

1. Tahap tokenizing, data teks artikel dibersihkan dari karakter yang mengandung selain huruf abjad (A-Z) kemudian dipecah menjadi array kata.
2. Tahap token filtering, array kata disasring dengan menghilangkan stopword
3. Kemudian pada tahap stemming & lemma, array kata artikel ditransformasikan menjadi kata dasar dengan menghilangkan imbuhan dan menyempurnakan kata agar sesuai dengan kata dasar.
4. Kata dasar yang dihasilkan kemudian diproses pada tahap word bigram untuk merangkai kata menjadi sebuah frase dengan komposisi 2 kata tanpa menghilangkan kata dasar sebelumnya.
5. Kemudian pada tahap term frequency, data frase kata dan data kata dasar ditransformasi menjadi tabel yang memuat kata dan jumlah (frekuensi) kata dalam satu berita.
6. Setelah itu, pada tahap data filtering, tabel term frequency disaring menjadi keyword. Penentuan keyword dipilih dari kata yang memiliki jumlah (frekuensi) yang melebihi threshold. Penentuan threshold (t) didapatkan dari jumlah TF maksimal (max\_*TF)* dibagi dua.

Pengujian pertama dari Data Processing adalah tahap tokenizing.

Tahap Tokenizing menghilangkan karakter yang menghilangkan karakter yang mengandung selain huruf (A-Z) dan kemudian dipecah menjadi array kata seperti yang ditunjukkan pada Tabel 4.2. Tahap ini teks dipecah menjadi array kata berdasarkan spasi.

### 4.6.2 Eksperimen Text Mining

Tabel 4.2 Tabel Data Pengujian Case Folding

|  |  |
| --- | --- |
| Input | Output |
| Bagaimana pembayaran untuk mahasiswa yang berbeasiswa Bidikmisi? | bagaimana pembayaran untuk mahasiswa yang berbeasiswa bidikmisi? |
| Kegiatan UKM apa saja yang ada di PENS? | kegiatan ukm apa saja yang ada di pens? |
| Beasiswa apa saja yang bekerja sama dengan PENS? | beasiswa apa saja yang bekerja sama dengan pens? |

Dari tabel 4.2 dapat dilihat bahwa proses case folding dapat memberikan hasil sesuai ekspektasi dimana semua karakter alphabet dirubah menjadi lowercase.

Tabel 4.3 Sampel Data Pengujian Remove Punctuation

|  |  |
| --- | --- |
| Input | Output |
| bagaimana pembayaran untuk mahasiswa yang berbeasiswa bidikmisi ? | bagaimana pembayaran untuk mahasiswa yang berbeasiswa bidikmisi |
| kegiatan ukm apa saja yang ada di pens? | kegiatan ukm apa saja yang ada di pens |
| beasiswa apa saja yang bekerja sama dengan pens? | beasiswa apa saja yang bekerja sama dengan pens |

Dari tabel 4.3 dapat dlihat bahwa proses remove punctutation dapat memberikan hasil sesuai ekspektasi dimana semua karakter symbol selain alphabet dan angka akan dihapus.

Tabel 4.4 Sampel Data Pengujian Stripping

|  |  |
| --- | --- |
| Input | Output |
| Bagaimana pembayaran untuk mahasiswa yang berbeasiswa bidikmisis | bagaimana pembayaran untuk mahasiswa yang berbeasiswa bidikmisi |
| kegiatan ukm apa saja yang ada di pensa | kegiatan ukm apa saja yang ada di pens |
| beasiswa apa saja yang bekerja sama dengan pensa | beasiswa apa saja yang bekerja sama dengan pens |

Dari tabel 4.4 dapat dilihat bahwa proses stripping dapat memberikan hasil sesuai ekspektasi dimana semua karakter whitespace berlebih di antara dua kata, di awal teks dan di akhir teks terhapus.

Tabel 4.4 Sampel Data Tokenizing

|  |  |
| --- | --- |
| Input | Output |
| bagaimana pembayaran untuk mahasiswa yang berbeasiswa bidikmisi | ['bagaimana', 'pembayaran', 'untuk', 'mahasiswa', 'yang', 'berbeasiswa', 'bidikmisi'] |
| kegiatan ukm apa saja yang ada di pens | ['kegiatan', 'ukm', 'apa', 'saja', 'yang', 'ada', 'di', 'pens'] |
| beasiswa apa saja yang bekerja sama dengan pens | ['beasiswa', 'apa', 'saja', 'yang', 'bekerja', 'sama', 'dengan', 'pens'] |

Dari tabel 4.4 dapat dilihat bahwa proses Tokenizing dapat memberikan hasil sesuai ekspektasi dimana kata-kata yang berada dalam daftar stopword akan dihapuskan.

Tabel 4.5 Sampel Data Filtering

|  |  |
| --- | --- |
| Input | Output |
| ['bagaimana', 'pembayaran', 'untuk', 'mahasiswa', 'yang', 'berbeasiswa', 'bidikmisi'] | ['pembayaran', 'mahasiswa', 'berbeasiswa', 'bidikmisi'] |
| ['kegiatan', 'ukm', 'apa', 'saja', 'yang', 'ada', 'di', 'pens'] | ['kegiatan', 'ukm', 'pens'] |
| ['beasiswa', 'apa', 'saja', 'yang', 'bekerja', 'sama', 'dengan', 'pens'] | ['beasiswa', 'pens'] |

Dari tabel 4.5 dapat dilihat bahwa proses filtering dapat memberikan hasil sesuai ekspektasi dimana kata-kata yang berada dalam daftar stopword akan dihapuskan.

Tabel 4.6 Sampel Data Stemming

|  |  |
| --- | --- |
| Input | Output |
| ['pembayaran', 'mahasiswa', 'berbeasiswa', 'bidikmisi'] | ['bayar', 'mahasiswa', 'beasiswa', 'bidikmisi'] |
| ['kegiatan', 'ukm', 'pens'] | ['giat', 'ukm', 'pens'] |
| ['beasiswa', 'pens'] | ['beasiswa', 'pens'] |

Dari tabel 4.6 dapat dilihat bahwa proses data preprocessing memberikan hasil sesuai ekspektasi setelah dilakukan tokenizing, filtering, dan stemming

Tabel 4.7 Summary Text Mining

|  |  |
| --- | --- |
| Input | Output |
| Bagaimana pembayaran untuk mahasiswa yang berbeasiswa Bidikmisi? | bayar mahasiswa beasiswa bidikmisi |
| Kegiatan UKM apa saja yang ada di PENS? | giat ukm pens |
| Beasiswa apa saja yang bekerja sama dengan PENS? | beasiswa pens |

Dari tabel 4.7 dapat dilihat bahwa proses text mining memberikan hasil sesuai ekspektasi setelah dilakukan tokenizing, filtering, dan stemming

### 4.6.3 Eksperimen Text Vectorizing (TF-IDF)

Pengujian Text Vectorizing adalah pengujian yang dilakukan untuk mengetahui apakah sistem berhasil mengubah teks input menjadi vektor bilangan desimal.

Tabel 4.8 Eksperimen Text Vectorize

|  |  |
| --- | --- |
| Input | Vektor |
| bayar mahasiswa beasiswa bidikmisi | [0.4471514625237128, 0.4471514625237128, 0.388785288863957, 0.6700426391572246] |
| giat ukm pens | [0.6102920003438043,  0.6957051393190256,  0.37886413586067147] |
| beasiswa pens | [0.7291602006183961, 0.6843430439729334] |

Dari tabel 4.8, terlihat bahwa proses vektorisasi menggunakan metode TF-IDF telah berjalan dengan baik. Hal ini terbukti dari keberhasilan sistem dalam membentuk vektor teks input yang diberikan. Proses konversi data input yang bertipe string menjadi bilangan desimal dilakukan dengan langkah-langkah berikut:

1. **Term Frequency (TF)**

Untuk seetiap kata dalam teks input, hitugn berapa kali kata tersebut muncul didalamnya. Ini akan memberi kita nilai TF untuk setiap kata dalam dokumen. Biasanya, TF dinormalisasi dengan membaginya dengan jumlah total kata dalam dokumen untuk menghindari bias terhadap dokumen yang lebih panjang. Ini disebut sebagai TF Normalized.

1. **Inverse Document Frequency (IDF)**

Hitung jumlah total dokumen dalam korpus. Untuk setiap kata dalam teks input, hitung berapa banyak dokumen dalam korpus yang mengandung kata tersebut. Hitung nilai IDF untuk setiap kata dengan menggunakan rumus IDF = log(total dokumen / jumlah dokumen yang mengandung kata tersebut).

1. **TF-IDF Score**

Kalikan nilai TF dengan nilai IDF untuk setiap kata dalam teks input. Ini akan memberi kita nilai TF-IDF untuk setiap kata dalam dokumen. Ini menghasilkan matriks TF-IDF, dimana setiap baris mewakili dokumen dan setiap kolom mewakili kata dalam kosakata, dan entri matriks tersebut adalah skor TF-IDF dari kata tersebut dalam dokumen tersebut.

### 4.6.4 Eksperimen Cosine Similiarity

Pengujian cosine similiarity adalah pengujian yang dilakukan untuk mengetahui apakah sistem berhasil melakukan pencocokan antara 2 vektor yang ada.

|  |  |
| --- | --- |
| INPUT | OUTPUT |
|  |  |
|  |  |
|  |  |
|  |  |

Dari Tabel 4.XX terlihat bahwa cosine similiarity melakukan pencocokan antara input pengguna dengan data yang ada dalam dataset. Pada eksperimen ini, saya menggunakan dataset yang terdiri dari total 15 baris data untuk memudahkan pengamatan selama eksperimen. Output yang dihasilkan ketidakcocokan total, sementara angka 1 mengindikasikan kecocokan total antara hasil input pengguna dan dalam dataset. Posisi nilai tersebut juga mencerminkan indeks data didalam dataset. Sebagai contoh, pada eksperimen pertama, data terletak pada indeks 0, pada eksperimen kedua, data terletak pada indeks 5, dan pada eksperimen ketiga, data terletak pada indeks 5, dan pada eksperimen ketiga, data terlektak pada indeks 10. Proses pencarian jawaban dengan cosine similiarity dilakukan dengan langkah-langkah sebagai berikut:

1. **Pemahaman Representasi Teks**

Teks input telah diubah menjadi representasi numerik mengganti TF-IDF. Setiap dokumen direpresentasikan sebagai vektor di dalam ruang kosinus, dimana setiap diimensi vektor mewakili skor TF-IDF dari kata-kata dalam korpus.

1. Perhitungan Cosine Similiarity

Pada langkah ini, dilakukan perhitungan cosine similiarity

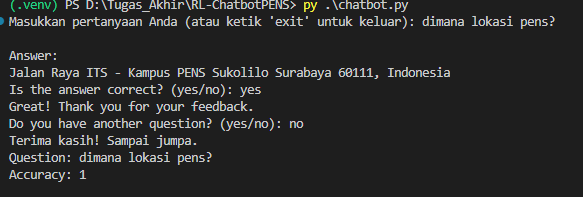
Antara dua vektor, yatu vektor yang mewakili teks input dan vektor yang mewakili dataset yang telah ada sebelumnya.

1. Interpretasi Hasil

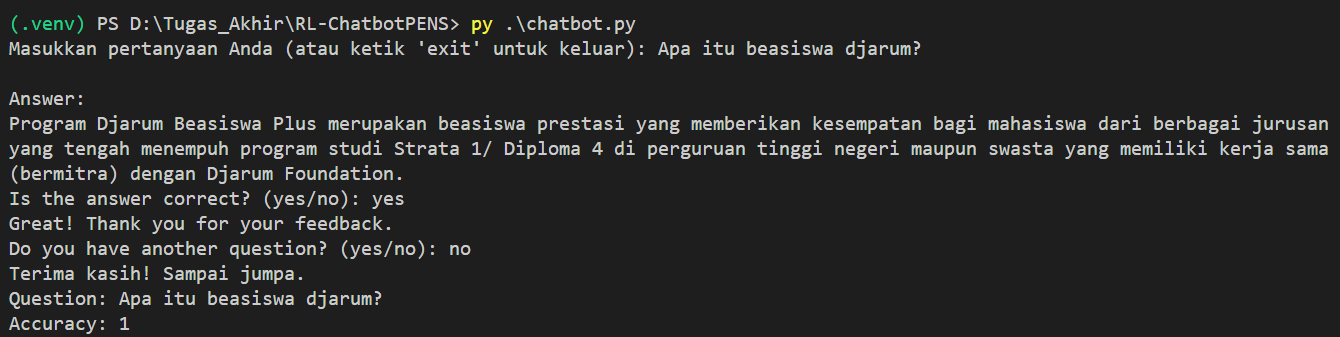
Hasil cosine similiartity adalah angka antara 0 dan 1, dimana nilai 1 menunjukkan bahwa dua vektor identik, dan nilai 0 menunjukkan bahwa dua vektor memiliki ketidakcocokan total. Semakin tinggi nilai cosine similiartity dantara dua vektor, semakin serupa atau mendekati nilai vektor-vektor tersebut.

### 4.6.5 Check Synonym

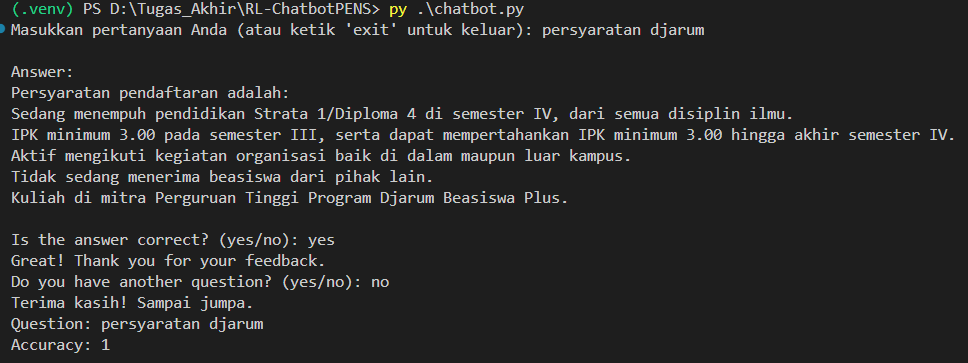
## 4.7 Pengujian Pencarian Jawaban

****

Gambar 4.1 Hasil Eksperimen Pencarian Jawaban 1

****

Gambar 4.2 Hasil Eksperimen Pencarian Jawaban 3



Gambar 4.3 Hasil Eksperimen Pencarian Jawaban 4

## 4.8 Pengujian Reinforcement Learning

### 4.8.1 Pengujian Pembuatan Matriks Reinforcement Learning

Pengujian Knowledge Reinforcement Learning merupakan percobaan dalam pembentukan satu matriks besar yang menampung seluruh keyword dari semua knowledge. Tabel 4.XXX menjelaskan mengenai detail dari pengujian Knowledge Reinforcement Learning

|  |  |
| --- | --- |
| Nama | Matriks Reinforcement Learning |
| Tujuan |  |
| Peralatan |  |
| Waktu Pengujian |  |
| Parameter Pengujian |  |

Pengujian pada tahap pembuatan Reinforcement Learning terdiri dari beberapa tahap, seperti yang ditunjukan pada skema berikut.

1. Mengambil keyword dari masing-masing knowledge kemudian mentransformasi menjadi kata dan frekuensi dari kata tersebut.
2. Membentuk matriks besar menggunakan kata dan id knowledge

Pengujian pertama adalah mengambil fitur dari masing-masing knowledge, sehingga masing-masing knowledge diambil dan dijadikan sebagai kolom fitur dari matriks besar yang ditunjukkan pada Tabel 4.XX, Kemudian id knowledge dijadikan sebagai baris matriks besar tersebut, dan matriks tersebut berisi frekuensi kata dari masing-masing artikel seperti yang ditunjukkan pada Tabel 4.XX

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Fitur-1 | Fitur-2 | Fitur-3 | Fitur-4 | Fitur-5 | Fitur-n |
| Knowledge-1 |  |  |  |  |  |  |
| Knowledge-2 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Setelah itu percobaan yang terakhir adalah membuat matris besar yang membuat fitur dari setiap knowledge. Pembuatan matriks ini memerlukan pertimbangan yang matang, karena matriks yang dibuat harus dapat bertambah sesuai dengan jumlah kata dan knowledge yang masuk. Kemudian, pertumbuhan datanya bersifat dua dimensi, dapat bertumbuh secara vertical dan horizontal.

#### 4.8.1.1 Pengujian Matriks Reinforcement Learning

Pengujian knowledge Reinforcement Learning merupakan percobaan untuk menghitung nilai kemiripan antar query dan matriks knowledge yang disediakan

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Pengujian pada tahap knowledge Reinforcement Learning terdiri dari beberapa tahap, seperti yang ditunjukkan pada skema berikut.

1. Melakukan Data Preprocessing terhadap query
2. Mengambil keyword dari masing-masing query kemudian mentransformasikan menjadi kata dan frekuensi dari kata tersebut
3. Mencari fitur dari matriks knowledge dan mengambil nilai fitur paling tertinggi.

Untuk tahap pertama sama seperti pada Text Processing sebelumnya, Data Preprocessing-nya terdapat pada query. Setelah query tersebut melalui Data Preprocessing, query akan bertransformasi menjadi sebuah kata dan frekuensi dari setiap kata tersebut.

**Tabel 4.X Hasil Tokenizing Query**

|  |
| --- |
| **Raw Text** |
| Apa itu Politeknik Elektronika Negeri Surabaya? |
| **Hasil Tokenize** |
| ['apa', 'itu', 'politeknik', 'elektronika', 'negeri', 'surabaya'] |

**Tabel 4.X Hasil Tokenizing, Stemming, Lemma Query**

|  |
| --- |
| **Hasil Tokenize** |
| Apa itu Politeknik Elektronika Negeri Surabaya? |
| **Hasil Filtering & Stemming Lemma** |
| ???? |

**Tabel 4.X Hasil Tokenizing, Stemming, Lemma Query**

|  |  |
| --- | --- |
| **Query** | **Frequency** |
| Politeknik | 1 |

Setelah menjadi per kata dan frekuensi, maka query akan diproses kemudian mencari apakah query tersebut ada di kolom fitur yang ebrada di dalam matriks *knowledge*.

**Tabel 4.X Matriks Knowledge untuk Querys**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Pada Tabel 4.XX merupakan contoh pengecekan terhadap *query* yang menmukan sebuah fitur didalam matriks knowledge. Kemudian ketika sudah menemukan fiturnya, maka akan dicek manakah nilai tertinggi yang terkandung di dalam fitur ‘PENS’ dan akan secara otomatis akan mengarah ke tabel knowledge yang mempunyai nilai fitur yang paling tinggi.

**Tabel 4.X Fitur yang diarahkan ke Knowledge**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Pada tabel 4.xx bisa dilihat bahwa nilai fitur yang paling tertinggi berada di Knowledge ke – 2 Untuk lebih jelasnya isi knowledge seperti apa, berikut tabel matriks knowledge yang ditunjukkan pada Tabel 4.XX

**Tabel 4.X Isi Kolom Knowledges**

|  |  |  |  |
| --- | --- | --- | --- |
| Knowledge | Fitur-1 | Depresi | Fitur N |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Kemudian, setelah mendapatkan nilai fitur yang paling tertinggi dan mengarahkan ke knowledge yang dituju, maka akan mengeluarkan sebuah output dari knowledge

**Gambar 4.2 Hasil Jawaban Chatbot**

Gambar 4.2 adalah hasil knowledge yang dituju. Pengguna menanyakan “Apa itu PENS?”. Sistem chatbot akan membaca input pengguna dan diproses didalam matriks knowledge. Sehingga akan mengeluarkan output dari knowledge yang dituju.

Setelah mendapatkan hasil output yang didapatkan dari matriks knowledge, maka ada sebuah pilihan yang mengharuskan pengguna unutk memilih ‘Ya’ atau ‘Tidak’. Ketika pengguna melihat jawaban dari chatbot tersebut dirasa sudah benar, maka pengguna bisa memilih opsi ‘Ya’ pada chatbot tersebut.

**Gambar 4.3 Pengguna memilih Jawaban Ya**

Kemudian, setelah pengguna memilih jawaban ‘Ya’, maka nilai fitur tersebut akan diambil dan diproses dengan metode Reinforcement Learning.

Pernyataan diatas merupakan rumus metode Reinforcement Learning dengan keterangan :

= nilai fitur di dalam matriks knowledges

Untuk perhitungan rumus tersebut, nilai didapatkan dari nilai fitur yang bernilai 3, karena nilai tersebut merupakan nilai yang paling besar di dalam fitur ‘PENS’. Maka, hasil perthitungan tersebut adalah : 3.0 + (0.1 \* 3.0) = 3.3

Setelah menghitung metode yang dihasilkan, nilai 3.3 akan update ke dalam nilai 3.0 nilai tersebut akan berubah yang awalnya 3 menjadi 3.3

**Tabel 4.XX Hasil Update Query**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Berlaku juga ketika pengguna menanyakan lagi pertanyaan yang sama “Apa itu PENS?” akan ada proses text processing lagi.

#### 4.8.1.2 Pengujian Kedua

**Tabel 4.XX Matriks Knowledge untuk Query Kedua**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Pada tabel 4.XX bisa dilihat bahwa fitur nilai yang paling tertinggi berada di Knowledge ke -2. Untuk lebih jelasnya isi knowledge nya seperti apa, berikut tabel matriks knowledge yang ditunjukkan pada Tabel balala.

**Tabel 4.XX Matriks Knowledge untuk Query Kedua**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Kemudian, setelah mendapatkan nilai fitur yang paling tertinggi dan mengarahkan ke knowledge yang dituju, maka akan mengeluarkan output dari knowledge.

**Gambar 4.xx Hasil Jawaban Chatbot Kedua**

Gambar 4.5 adalah hasil knowledge yang dituju. Pengguna menanyakan “Apa itu PENS?”. Sistem chatbot akan membaca input pengguna dan diproses didalam matriks knowledges. Sehingga akan mengeluarkan output dari *knowledge* yang dituju.

Setelah mendapat hasil output yang didapatkan dari matriks knowledge, maka ada sebuah pilihan yang mengharuskan pengguna untuk memilih 'Ya’ atau ‘Tidak’. Untuk percobaan ini, pengguna mencoba untuk memilih jawaban ‘Tidak’ pada chatbot Tersebut.

**Gambar 4.xx Pengguna memilih jawaban Tidak**

Kemudian, setelah pengguna memilih jawaban “Tidak”, maka nilai fitur tersebut akan diambil dan diproses dengan metode *Reinforcement Learning.*

Pernyataan diatas merupakan metode Reinforcement Learning dengan keterangan

α = nilai fitur dialam matriks Knowledge.

Untuk perhitungan rumus tersebut, nilai α didapatkan dari nilai fitur yang bernilai 3.3, karena nilai tersebut merupakan nilai yang paling besar didalam fitur “PENS”. Maka, hasil perhitungan tersebut adalah : 3.3 – (0.1 \* 3.3) = 3.0

Pengujian selanjutnya adalah mencoba untuk mengganti jawaban yang disediakan dari chatbot. Karena system jawaban knowledge berupa ranking

# BAB 5

**PROGRES PENELITIAN**

**5.1 Hal yang sudah dikerjakan**

Sejalan dengan permasalahan yang ditemukan, Peneliti sudah berhasil mencover mulai dari pengerjaan Model hingga selesai.

**5.2 Hal yang belum dikerjakan**

* Proses Deployment pada Chatbot Reinforcement Learning
* Uji Testing Kepada Real User untuk mendapatkan Akurasi tertinggi
* Fixing Bug pada pertanyaan yang belum sesuai dengan jawaban.

**5.3 Kendala**

Disini berikan penjelasan terkait dengan kendala-kendala yang dihadapi oleh sewaktu melakukan penelitian

1. Logika integrasi dari model ke website yang membutuhkan waktu untuk belajar.

# Daftar Pustaka

[1] Desy Nur Pratiwi dkk.(2022). Financial Technology (FINTECH): Generasi Z dan Generasi Milenial,4.

[2] Fahmi Ahmad Burhan (31 Mei 2022). Teknologi Chatbot Digandrungi, Pengguna Startup Kata.ai Tumbuh 170%.

Diakses dari <https://katadata.co.id/yuliawati/digital/6295e9c6832d5/teknologi-chatbot-digandrungi-pengguna-startup-kataai-tumbuh-170> (accessed Jun 7, 2023.

[3] Zifora Nur Baiti 2013, Fresy Nugroho, ST.,MT. “Aplikasi Chatbot”M13” Untuk Informasi jurusan Teknik Informatika Berbasis Sistem Pakar Menggunakan Metode Forward Chaining”

DOI:[10.18860/mat.v0i0.2425](http://dx.doi.org/10.18860/mat.v0i0.2425)

[4] IBM – “ What Is Text Mining ”

Diakses dari : https://www.ibm.com/topics/text-mining#:~:text=the%20next%20step-,What%20is%20text%20mining%3F,meaningful%20patterns%20and%20new%20insights. (accessed Jun 10 7, 2023)

[5] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, Martin Riedmiller 2013, “Playing Atari with Deep Reinforcement Learning”

[6] Ahmad Hammoudeh, “A Concise Introduction to Reinforcement Learning” DOI: 10.13140/RG.2.2.31027.53285,2018.

<https://mentormate.com/blog/what-is-heroku-used-for-cloud-development/> (Accessed June 15, 2023)

[9] Rifki Achsani Dwi Taqwim(2020) “Chatbot Gangguan Mental Berbasis Incremental Knowledge dengan Metode Differential Learning”.

[10] Ariata.C(2023) Ap aitu MySQL? “Pengertian MySQL, Cara Kerja dan Kelebihannya” <https://www.hostinger.co.id/tutorial/apa-itu-mysql> (accessed June 21,2023)

[11] Dwi Septiani, Ica Isabela “ Analisis Term Frequency Document Frequency (TF-IDF) Dalam Temu Kembali Informasi Pada Dokumen Teks

[12] Keelan Graham-Sands “Dissertation Deep Reinforcement Learning for Dialog-Based System”

[13] Bram Bakker. Reinforcement Learning with Long Short-Term Memory, Dept. of Psychology, Leiden University / IDSIA

[14] Jiwei Li, Alan Ritter, Michel Galley, Jianfeng Gao, Dan Jurafsky. “Deep Reinforcement Learning for Dialogue Generation. Stanford University

[15] Ivan Aulia Rahman(20230 “Chatbot Layanan Informasi PENS dengan Context Recognition

[16] D. Kurniadi and A. Mulyani, “Implementasi Pengembangan Student Information Terminal (S-IT) Untuk Pelayanan Akademik Mahasiswa,” Jurnal Algoritma, vol. 13, no. 2, pp. 437–442, 2017, doi: 10.33364/algoritma/v.13-2.437.

[17] E. Mursidah, L. Ambarwati, A. Karima, K. Kunci, : Chatbot, and M. Baru, “Implementasi Chatbot Layanan Informasi Pendaftaran Mahasiswa Baru Program Pascasarjana Departemen Teknik Informatika Its,” Jurnal Ilmiah NERO, vol. 7, no. 1, p. 2022, 2022.