

2. LANDASAN TEORI

2.1. Tinjauan Pustaka

2.1.1. Chatbot

Chatbot merupakan program komputer yang didesain untuk mensimulasikan pembicaraan dengan sesama pengguna manusia melalui teks/suara. Saat ini chatbot telah digunakan dalam berbagai bidang seperti kesehatan, bisnis, dan edukasi (Kuhail et.al, 2022). Berdasarkan cara chatbot mengelola *input* dan memberi respons balik, chatbot dapat terbagi menjadi tiga, rule-based model, retrieval model, dan generative model. (Adamopoulou & Moussiades, 2020)

2.1.2. Rule-Based Chatbot

Rule based model merupakan chatbot yang menggunakan *pre-determined rules* (pertanyaan dan jawaban) untuk menentukan respons yang akan diberikan ke *user* (Adamopoulou & Moussiades, 2020). Namun, jika input *user* tidak tercantum pada rules yang ada maka model ini tidak dapat memberikan respons seperti yang diharapkan. Model ini juga memiliki kekurangan dalam menangani kesalahan pengejaan dan tata bahasa pada input *user*. Rule-Based model bagus digunakan untuk mengatasi pertanyaan yang mudah dan memiliki jawaban yang jelas, namun diperlukan penanganan lebih lanjut untuk mengatasi pertanyaan yang lebih kompleks. (Thorat et.al, 2020)

2.1.3. Scripted Chatbot

Scripted Chatbot merupakan chatbot yang menerima input dari *user* berupa *specific command/query* secara langsung sehingga chatbot dapat memberikan jawaban sesuai dengan *script* yang telah disiapkan sebelumnya. Adapun kekurangan dari chatbot jenis ini adalah kurangnya fleksibilitas terhadap input yang dapat *user* berikan, sehingga menghasilkan percakapan yang cenderung kaku dan tidak natural.

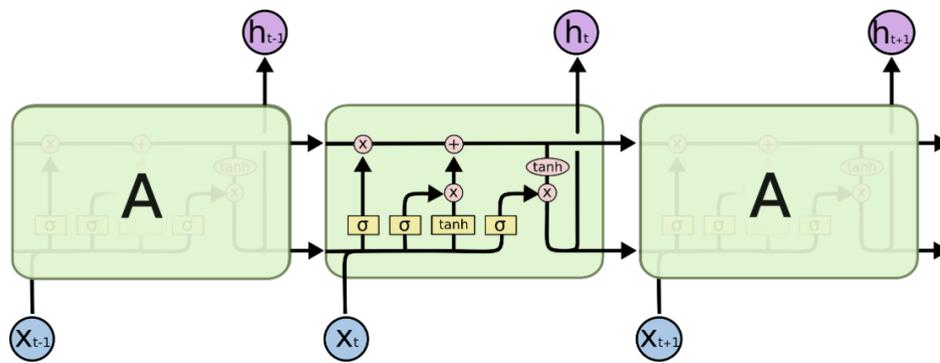
2.1.4. Intent Classification

Intent classification merupakan sebuah proses yang menjadi salah satu *core feature* dari chatbot. *Intent classification* berfungsi untuk mengklasifikasikan input

dari *user* ke dalam beberapa kategori untuk menghasilkan jawaban yang sesuai dengan keinginan *user*. Beberapa metode yang sering digunakan diantaranya adalah Convolutional Neural Network (CNN), Recurrent Neural Network (RNN)(Zhang & LeCun, 2015), Long Short Term Memory (LSTM)(Gennaro et.al, 2020), dan API seperti DialogFlow(Chandra et al., 2020).

2.1.5. Long Short Term Memory (LSTM)

LSTM merupakan salah satu jenis RNN yang telah dikembangkan untuk mengatasi masalah dependensi jangka panjang pada RNN. Salah satu perbedaan LSTM dibandingkan RNN biasanya adalah terdapat 4 neural network layer (3 sigmoid, 1 tanh) di setiap modul LSTM. Berikut pada Gambar 2.1 merupakan gambar visualisasi dari model LSTM.



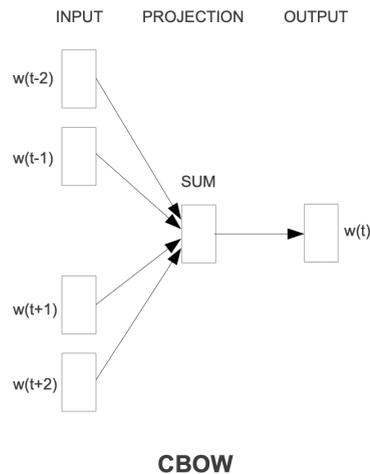
The repeating module in an LSTM contains four interacting layers.

Gambar 2.1 Visualisasi LSTM

Sumber: (*Understanding LSTM Networks -- Colah's Blog*, n.d.)

2.1.6. Word2Vec

Word2Vec merupakan salah satu teknik embedding yang banyak digunakan untuk merepresentasikan kata menjadi vector. Skripsi ini akan mengimplementasikan penggunaan Word2Vec model CBOW (*Continuous bag of Words*) yang berguna untuk memprediksi kata berdasarkan kata di sekitarnya (Mikolov, 2013).



Gambar 2.2 Perbandingan model CBOW dan Skip-gram

Sumber: Mikolov et al. (2013)

2.1.7. Word2Vec Metode Dieka Nugraha (2018)

Word2Vec versi Dieka Nugraha ini menggunakan dataset berupa hasil dump wikipedia (idwiki-latest-pages-articles.xml.bz2) pada tahun 2020 sebagai corpus. Dataset pada file bz2 tersebut kemudian akan diekstrak dan melalui preprocess sederhana tanpa lemmatization menggunakan WikiCorpus. Data yang telah diekstrak lalu disimpan ke dalam sebuah file.

Modelnya sendiri di train dengan menggunakan gensim dengan dimensi 200, dan menggunakan function LineSentence untuk menghasilkan generator yang membaca file text dari data yang telah diekstrak. Hasil pengujian analogy menggunakan laki : raja, maka perempuan : ? memberi hasil 0.914 pada kata 'ratu'.

2.1.8. Word2Vec Metode Eko Setiawan (2023)

Word2Vec versi Eko Setiawan menggunakan dataset hasil scrape WikiHow. Data ini terdiri dari 16.308 artikel dari WikiHow, masing-masing artikel terdiri dari beberapa langkah yang berisi teks penjelasan dalam bahasa indonesia. Dikarenakan tiap artikel memiliki beberapa bagian terpisah seperti: link artikel berbahasa indonesia, judul dari tiap metode, kesimpulan dari tiap metode, ringkasan langkah-langkah dari tiap metode, judul dalam bahasa inggris dan link artikel dalam bahasa inggris. Oleh karena itu diperlukan langkah lanjutan sebelum data tersebut dapat dipakai sebagai corpus. Eko Setiawan sendiri memutuskan untuk hanya

menggunakan bagian judul dari masing-masing metode (*sub_title*), kesimpulan dari metode (*summary*) dan ringkasan langkah-langkah (*document*) metode tersebut. Dan menampilkannya menggunakan *dataframe*. Lalu Eko Setiawan menggabungkan kesimpulan dari tiap metode (*summary*) dengan ringkasan langkah-langkah (*document*) untuk memperbanyak kosa kata dataset yang akan digunakan.

Teks yang telah digabungkan lalu akan melalui tahap *preprocessing* berupa *case folding*, *punctuation removal*, *new line removal*, *text extraction* (hanya mengambil *text* berupa huruf dan spasi), sebelum di tokenisasi menggunakan *function tokenize* yang telah disediakan oleh library *gensim*. Data yang telah ditokenisasi lalu akan digunakan untuk melatih model *Word2Vec*.

Model *Word2Vec*nya sendiri menggunakan parameter berupa *window=10*, *min_count=5*, *workers=4*, *epochs=10*. Model ini menghasilkan akurasi sebesar 0.715 pada *similarity* antara kata 'cerdas' dan 'pintar'. Selain itu model ini juga menghasilkan akurasi sebesar 0.697 untuk kata 'ratu' pada tes analogi laki : raja, maka perempuan : ?.

2.1.9. GloVe

GloVe merupakan model *embedding* kata yang merepresentasikan kata dalam bentuk vektor dengan menganalisa statistik dari frekuensi kata yang muncul secara bersamaan. Salah satu kelebihan menggunakan GloVe adalah model ini tidak hanya menganalisa hubungan kemunculan antar teks, tapi juga dapat menganalisa hubungan semantik antar teks, seperti analogi dan kemiripan. (Pennington & Manning, 2014)

2.1.10. A/B/C Testing

A/B testing atau yang biasa disebut dengan A/B/n testing merupakan sebuah metode untuk membandingkan performa dari dua atau lebih strategi secara objektif untuk mendapatkan strategi terbaik. A/B/C testing merupakan salah satu bentuk lanjutan dari A/B testing. Test ini dijalankan dengan memberi 3 versi berbeda dari sebuah objek test kepada beberapa grup *user* secara bersamaan, lalu mengambil dan menganalisa data yang didapatkan dari grup-grup tersebut dengan perhitungan statistik dan mengambil hasil terbaik. (Sheng et.al, 2023)

2.2. Tinjauan Studi

2.2.1. Sistem Chatbot dengan Metode Naive Bayes Untuk Menanggapi Pertanyaan Mahasiswa Tentang Pra Skripsi (Giovanni, 2023)

Masalah yang diangkat dalam penelitian ini adalah keterbatasan dosen pembimbing untuk melayani pertanyaan mahasiswa mengenai seputar pra-skripsi seperti judul dan problem skripsi yang cenderung repetitif.

Metode yang diusulkan oleh penelitian ini adalah Naive Bayes untuk klasifikasi, K-Means untuk *clustering*, Doc2Vec, Word2Vec, dan Cosine Similarity.

Hasil dari penelitian ini sudah cukup baik, yaitu accuracy sebanyak 72% dalam memberikan jawaban yang benar pada tiap pertanyaan dengan 5 *test data*.

Walaupun memiliki tujuan yang serupa, skripsi ini memiliki perbedaan dengan penelitian ini pada metode yang digunakan. Penelitian ini menggunakan metode Naive Bayes untuk klasifikasi pertanyaan dan K-Means untuk *clustering* data. Sedangkan skripsi ini akan menggunakan pendekatan Rule-Based Method sebagai *outputnya*, dan menggunakan LSTM dengan Pre-trained model GloVe dan Word2Vec untuk mengklasifikasikan *intent* / maksud dari *input user*.

2.2.2. Intent Classification in Question-Answering Using LSTM Architectures (Gennaro et.al, 2020)

Masalah yang diangkat pada penelitian ini adalah untuk menguji keefektifan metode *machine learning* pada chatbot untuk mengklasifikasikan *intent user* terhadap response yang sesuai.

Metode yang digunakan pada penelitian ini adalah LSTM menggunakan Pre-trained data GloVe dan dijalankan menggunakan algoritma BPTT untuk *intent recognition* dan *classification*, serta Bi-LSTM untuk *answer generation* dari model sebelumnya.

Penelitian ini membuktikan bahwa penggunaan LSTM dalam mengklasifikasikan intensi *user* sangat baik dengan akurasi pada *main category* sebesar 91,20% dan akurasi pada *sub-category* sebesar 82,20%.

Perbedaan dengan penelitian ini terletak pada hasil akhir yang dihasilkan, serta dataset yang digunakan. Walaupun menerapkan metode yang serupa, LSTM, namun dataset Pre-Trained yang digunakan dalam pembuatan model *intent classificationnya* memiliki perbedaan pada *size* data serta bahasa yang digunakan. Dataset yang

digunakan pada penelitian ini menggunakan 5500 pertanyaan sebagai *training data* dan 500 pertanyaan sebagai *test data*, dengan 6 *main category* dan masing-masing memiliki beberapa *sub-category*. Namun pada skripsi ini, nantinya dataset yang digunakan akan berupa kurang dari 1000 pertanyaan sebagai *training data*.

2.2.3. Implementasi Long Short-Term Memory pada Chatbot Informasi Akademik Teknik Informatika Universitas Lampung (Wintoro et.al, 2022)

Masalah yang diangkat pada penelitian ini adalah pemanfaatan chatbot sebagai media informasi akademik bagi mahasiswa Program Studi Teknik Informatika Universitas Lampung.

Metode yang digunakan pada penelitian ini adalah LSTM untuk mengklasifikasikan *input* berupa pertanyaan dari *user* ke dalam 29 jawaban yang telah disediakan.

Hasil dari penelitian ini sangat bagus dengan tingkat akurasi 99% dengan dataset uji 100 pertanyaan random dari 20 *user*.

Perbedaan penelitian ini terletak pada model LSTM yang digunakan. LSTM pada penelitian ini dikembangkan struktur model yaitu *Embedding layer*, LSTM, dan *Output layer*. Pada penelitian ini, dataset teks yang ada diubah ke dalam bentuk *sequence of integer* yang kemudian diubah menjadi vector sebelum diproses menggunakan LSTM. Sedangkan skripsi ini nantinya akan menggunakan GloVe dan Word2Vec sebagai Embedding Layer sebelum, di train lebih lanjut pada dataset yang tersedia lalu diproses menggunakan LSTM.

2.2.4. Rule-Based Chatbot for Student Enquiries (Singh et.al, 2019)

Masalah yang diangkat pada penelitian ini terletak pada banyaknya mahasiswa yang mengantri di layanan administrasi Asia Pasific University (APU). Hal ini menyebabkan terbuangnya banyak waktu mahasiswa sehingga kurangnya waktu mereka untuk mengejar prestasi akademik yang mereka inginkan.

Metode yang digunakan pada penelitian ini adalah pengumpulan data melalui interview kepada pihak administrasi, sekaligus melalui kuesioner yang dibagikan kepada mahasiswa untuk lebih mengetahui dan mengeksplor kebutuhan-kebutuhan yang diperlukan oleh mahasiswa. Data yang didapat dari interview kemudian dianalisa dan dikategorikan ke beberapa kelompok. Data ini juga akan dijadikan dasar

dari flowchart yang akan dijadikan fondasi dari program yang akan dibuat. Sedangkan data yang didapat dari kuesioner digunakan untuk menemukan korelasi dan hubungan dari data yang sudah ada yang memerlukan tindakan lebih lanjut.

Penelitian ini menyebutkan bahwa chatbot yang dibuat telah memenuhi kebutuhan *user* yang didapatkan dari data yang telah dikumpulkan. Namun penelitian ini memiliki kekurangan pada beberapa hal seperti aksesibilitas terhadap Backend chatbot kepada staff administrasi APU, hanya bisa menjawab pertanyaan pada tingkat informasi umum, dan *user* hanya bisa mengakses jawaban chatbot melalui *chat prompt/keyword* yang telah disediakan.

Perbedaan dengan penelitian ini terletak pada model chatbot yang akan digunakan. Chatbot yang akan dibuat pada skripsi ini akan terdiri dari 3 model, yang nantinya akan lebih diujikan kepada *user* sehingga menghasilkan chatbot yang paling optimal digunakan oleh *user*. Salah satu dari chatbot yang akan diujikan kepada *user* (mahasiswa Universitas Kristen Petra) dari skripsi ini juga merupakan Rule-Based Chatbot. Namun hasil yang didapatkan belum tentu sama dengan hasil dari penelitian ini dikarenakan perbedaan *user behavior* dari Asia Pasific University dengan Universitas Kristen Petra.