



Available online at : <http://bit.ly/InfoTekJar>

InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan

ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600



Implementasi Mesin Penerjemah Statistik Pada Aplikasi *Chatting* Berbasis Android Dengan Moses Decoder

Muhammad Nur Amin, Arif Bijaksana Putra Negara, Anggi Perwitasari

Jurusan Informatika, Fakultas Teknik, Universitas Tanjungpura, Jalan Prof. Dr. H. Hadari Nawawi, Pontianak, 78124, Indonesia

KEYWORDS

Aplikasi *Chatting*, Android, HMM, Moses Decoder, Mesin Penerjemah Statistik, Bahasa Indonesia, Bahasa Melayu Sambas

CORRESPONDENCE

Phone: +62 852 5221 5103

E-mail: muhammadamin0598@gmail.com

A B S T R A K

Chatting adalah salah satu bentuk komunikasi yang umumnya dilakukan antara dua orang atau lebih secara langsung (*realtime*) yang memanfaatkan fasilitas jaringan (LAN atau internet). Keberagaman bahasa yang luar biasa dan keterbatasan manusia dalam menguasai berbagai macam bahasa dapat menghambat tersampainya pesan atau informasi pada aktivitas *chatting*. Karakteristik bahasa *chatting* adalah menggunakan bahasa yang tidak baku sehingga muncul banyak sekali variasi dalam penuturan satu bahasa bergantung pada kondisi orang yang menuturkan. Tujuan dari penelitian ini adalah untuk membuat sebuah aplikasi *chatting* berbasis android yang mampu memperbaiki bahasa *chatting* ke bahasa baku dan menerjemahkan bahasa tersebut menggunakan mesin penerjemah statistik Moses Decoder, sehingga dapat digunakan pada komunikasi teks antar pengguna bahasa yang berbeda. Metodologi penelitian yang dilakukan meliputi analisis permasalahan, perancangan sistem, implementasi, dan pengujian sistem. Pada penelitian ini dilakukan pra proses pesan bahasa Melayu Sambas *chatting* dan penerjemahan bahasa Melayu Sambas ke bahasa Indonesia. Pengujian pada penelitian ini berupa pengujian akurasi hasil pra proses HMM, akurasi hasil terjemahan mesin penerjemah Moses Decoder, dan pengujian fungsionalitas aplikasi. Hasil pengujian pra proses HMM memperoleh nilai tertinggi 0,9 (*accuracy*), 0,834 (*precision*), 0,829 (*recall*), dan 0,832 (*f-measure*). Hasil pengujian akurasi terjemahan mesin translasi dengan BLEU memperoleh nilai tertinggi 53,14% sedangkan oleh ahli bahasa memperoleh nilai 86,86% dan 87,71%. Hasil pengujian *black box* menunjukkan aplikasi dapat mengirim, menerima, memperbaiki dan menerjemahkan pesan. Secara keseluruhan penelitian dinilai berhasil dalam membuat sebuah aplikasi *chatting* berbasis android yang dapat memperbaiki bahasa *chatting*, mengimplementasikan mesin penerjemah statistik Moses Decoder dan dapat digunakan sebagai media komunikasi teks.

PENDAHULUAN

Komunikasi ialah proses pertukaran pesan atau berita oleh dua orang atau lebih dengan maksud untuk memahami pesan. Bahasa adalah alat yang digunakan manusia dalam melakukan komunikasi, dengan bahasa seseorang dapat mengenalkan diri, menyampaikan ide, dan menceritakan pengalamannya kepada orang lain. Dalam pemakaiannya, keragaman bahasa dapat berupa bahasa lisan atau bahasa tulis tergantung dari kebutuhan dan tujuan komunikasi [1].

Berkembangnya teknologi informasi berdampak pada akses komunikasi yang menjadi lebih luas. *Chatting* merupakan bentuk komunikasi yang mana dilakukan secara langsung atau *realtime* di antara dua orang atau lebih dengan memanfaatkan fasilitas jaringan [2]. *Chatting* bukan lagi hal yang diminati, namun menjadi bagian dari kebutuhan sosial manusia untuk memudahkan komunikasi baik jauh maupun dekat [3]. Dalam <https://doi.org/10.30743/infotekjar.v6i1.4025>

melakukan komunikasi selama ini di kalangan masyarakat umum ketika berkomunikasi lewat *chatting* mengalami kendala ketika akan berkomunikasi dengan orang yang menggunakan bahasa yang berbeda [4].

Mesin penerjemah statistik merupakan mesin yang digunakan untuk melakukan proses penerjemahan bahasa dari satu bahasa ke bahasa lainnya secara otomatis. Hasil dari penerjemahan didapat atas dasar model statistik di mana parameter-parameternya berdasarkan hasil dari analisis korpus paralel (korpus teks bilingual), dan Moses Decoder merupakan salah satu sistem mesin penerjemah statistik yang sering digunakan [5].

Merupakan jenis komunikasi verbal dalam bentuk teks atau tulis, serta digunakan oleh berbagai lapisan masyarakat, umumnya bahasa dalam aktivitas *chatting* menggunakan kata tidak baku sehingga muncul variasi dalam penuturan bahasa oleh pelaku aktivitas *chatting*. Perubahan bentuk atau variasi bahasa pada

komunikasi teks dapat berupa abreviasi atau penyingkatan kata, penggunaan slang, maupun lain sebagainya.

Berdasarkan pembahasan yang telah dipaparkan, maka dilakukan penelitian untuk membuat implementasi dari mesin penerjemah statistik berupa aplikasi *chatting* berbasis android dengan memanfaatkan Moses Decoder sebagai mesin penerjemah dan HMM dengan *trigram* untuk perbaikan bahasa *chatting* ke bahasa baku, sehingga dapat digunakan sebagai media komunikasi tulis antar pengguna yang berbeda bahasa. Pada penelitian dilakukan penerjemahan bahasa Melayu Sambas ke bahasa Indonesia.

TINJAUAN PUSTAKA

Chatting

Komunikasi ialah proses pertukaran pesan atau berita oleh dua orang atau lebih agar pesan yang dimaksud dapat dipahami. Adanya layanan komunikasi dan informasi yang kompleks telah mendisrupsi kehidupan manusia masa kini dalam segala aspek, terutama dalam kehidupan sosial satu sama lain. Menjadi bagian dari media komunikasi pada bidang teknologi informasi, *chatting* dengan tujuan untuk menjalin hubungan dan interaksi sosial di dunia maya, merupakan salah satu dari layanan favorit hampir semua kalangan baik anak muda atau bahkan orang dewasa pada saat ini. Aktivitas *chatting* sendiri adalah komunikasi dengan bentuk yang berwujudkan bahasa tulis (tuturan verbal teks) dengan diikuti oleh penggunaan bentuk *emoticon* (simbol-simbol) atau kode-kode bahasa yang menjadi pengganti dari ujaran lisan yang karena keterbatasan media yang digunakan, ujaran tersebut tidak dapat diekspresikan ke dalam ungkapan verbal [6].

Merupakan jenis komunikasi verbal dalam bentuk teks atau tulis, serta digunakan oleh berbagai lapisan masyarakat dalam berbagai situasi dan konteks sosial, umumnya bahasa dalam aktivitas *chatting* menggunakan kata tidak baku yaitu kata yang digunakan pada situasi tidak resmi atau informal. Karena bahasa yang digunakan bersifat tidak baku maka muncul variasi dalam penuturan bahasa oleh pelaku aktivitas *chatting*. Faktor-faktor sosial yang memengaruhi variasi bahasa dapat diketahui berdasarkan latar belakang penutur bahasa, beberapa faktor tersebut seperti latar belakang pendidikan, usia, pekerjaan, gender, keadaan sosial ekonomi, tingkat kebangsawanan, dan sebagainya [7].

Normalisasi Teks

Dalam aktivitas *chatting* atau komunikasi teks penggunaan kata tidak baku menyebabkan perubahan bentuk pada suatu bahasa seperti terjadinya abreviasi atau penyingkatan kata, penggunaan slang, maupun lain sebagainya sehingga muncul banyak variasi dalam suatu bahasa. Tantangan terbesar dalam model komputasi bahasa diantaranya ambiguitas makna kata dan ambiguitas sintaksis, sementara penulisan atau ejaan yang tidak baku/sempurna merupakan salah satu penyebabnya [8]. Karena setiap orang dapat memiliki bentuk atau variasi yang berbeda-beda (yang tidak baku) dalam menuturkan satu bahasa pada komunikasi teks, maka diperlukan normalisasi teks.

Normalisasi teks atau normalisasi kata adalah pengubahan kata menjadi kata yang baku dalam suatu bahasa, dari kata yang tidak baku [9]. Normalisasi teks pada penelitian ini dengan *Hidden*

Markov Model (HMM) menggunakan rangkaian *trigram* (3 gram).

Hidden Markov Model (HMM)

HMM atau *Hidden Markov Model* adalah model statistik di mana pada suatu sistem yang dimodelkan diasumsikan dengan kondisi yang tidak dapat diamati (tidak terobservasi) sebagai *markov* proses. Suatu HMM bisa diumpamakan sebagai jaringan Bayesian dinamis yang sederhana (*simplest dynamic Bayesian network*) [10].

Pada model atau jenis *Markov* normal, pengamat dapat melihat semua keadaan secara langsung. Oleh karena itu, probabilitas kondisi transisi adalah satu-satunya dari parameter yang dapat diamati. Pada HMM keluaran yang memiliki ketergantungan terhadap keadaan dapat terlihat, sekalipun keadaan tidak terlihat secara langsung.

Setiap dari kondisi mempunyai porsi probabilitas pada semua kemungkinan keluar. Oleh sebab itu, tahapan dari langkah-langkah yang dihasilkan oleh HMM akan memberikan informasi tentang tahapan keadaan. Harus dapat dipahami bahwa untuk sifat tersembunyi (*hidden*) mengacu pada keadaan bagian yang dilalui model, bukan parameter model. Bahkan jika parameter model diketahui maka model tetap tersembunyi.

Hidden Markov Model bisa diterapkan atau diaplikasikan dalam bidang *temporal pattern recognition* "pengenalan pola temporal", contohnya pada pengenalan tulisan, gestur, suara, finansial, bioinformatika, kompresi kalimat, ekonomi, *computer vision*, dan pengenalan not balok. Pada penelitian ini model HMM digunakan untuk pra proses atau perbaikan kalimat *chatting* bahasa Melayu Sambas menjadi kalimat baku bahasa Melayu Sambas.

N-gram

Pemodelan *N-gram* pada dasarnya ialah model probabilistik, yang selanjutnya dikembangkan dengan tujuan mencari kemungkinan *item* berikutnya pada urutan *item* yang diberikan. Tergantung pada aplikasinya, *item* bisa berbentuk kata, huruf/karakter, atau bentuk lainnya. Satu dari model *N-gram* ialah model *N-gram* yang berbasis kata yang digunakan dalam memprediksi kemunculan kata berikutnya menurut urutan kata tertentu yang diberikan. Dalam artian tertentu, *N-gram* hanyalah kumpulan dari kata, dan setiap kata mempunyai panjang *n* kata. Misalnya, *N-gram* untuk ukuran 1 kata disebut "*unigram*"; "*bigram*" untuk ukuran 2 kata; "*trigram*" untuk ukuran 3 kata, dan begitu pula untuk seterusnya.

Dalam pembuatan kata, model *N-gram* dipakai dengan tujuan mengekstrak *n* fragmen kata dari rangkaian kata (kalimat, paragraf, bacaan) yang diberikan, yang dibaca secara terus-menerus dari teks sumber hingga akhir dokumen. Misalnya, pada kalimat: "Saya dapat melihat cahaya itu" bisa dipecah menjadi *N-gram* sebagai berikut:

- uni-gram: saya, dapat, melihat, cahaya, itu
- bi-gram: saya dapat, dapat melihat, itu ada
- tri-gram: saya dapat melihat, dapat melihat dia dan seterusnya.

A. *Natural Language Toolkit (NLTK)*

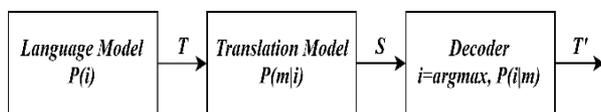
NLTK atau (*Natural Language Toolkit*) ialah *library*, modul atau serangkaian program yang dibangun untuk mempermudah dalam pengelolaan dan pemrosesan bahasa alami. NLTK untuk Bahasa Inggris awalnya dikembangkan pada tahun 2001 oleh Steven Bird dan Edward Loper dari Departemen Ilmu Komputer dan Informasi di Universitas Pennsylvania untuk mendukung penelitian dan pengajaran di NLP atau bidang terkait, seperti komputasi bahasa, kecerdasan buatan, pembelajaran mesin, dan lain-lain [11].

NLTK dibangun dengan bahasa pemrograman python dan pada penelitian ini digunakan untuk analisis teks yaitu pengujian *accuracy*, *precision*, *recall*, dan *f-measure* pada hasil pra proses HMM yang dikerjakan.

Mesin Penerjemah Statistik

Mesin penerjemah statistik ialah suatu mesin translasi di mana penerjemahan dijalankan dengan parameter-parameter yang diambil dari hasil analisis korpus paralel dengan berbasis model statistik. Konsep probabilitas ialah merupakan pendekatan statistik yang digunakan, di mana pada setiap pasangan kalimat (S,T) akan diberikan sebuah $P(T|S)$ yang didefinisikan sebagai distribusi probabilitas yang mana sebuah penerjemah ketika mendapatkan masukan S pada bahasa sumber maka akan memberikan hasil T pada bahasa target atau tujuan [12].

Pada mesin translasi berbasis statistik terdapat 3 komponen yaitu *language model*, *translation model*, dan *decoder*. Ketiga elemen/komponen inilah yang ikut serta dalam translasi bahasa dari satu bahasa ke bahasa lainnya atau dari bahasa sumber ke bahasa sasaran [13]. Ketiga komponen pada mesin penerjemah statistik diperlihatkan oleh Gambar 1 di bawah.



Gambar 1. Komponen Mesin Penerjemah Statistik

Pada penelitian ini mesin translasi dengan pendekatan statistik atau mesin penerjemah statistik dipilih sebagai alat penerjemahan bahasa. Penerjemahan dengan mesin penerjemah statistik memberikan hasil terjemahan yang lebih baik bila dibanding dengan hasil terjemahan kata per kata (*interlinear translation*), namun tentu harus memiliki kualitas korpus paralel (padanan kalimat-kalimat bahasa sumber dan bahasa tujuan) dengan kualitas baik dan cukup banyak jumlahnya yang dimasukkan ke dalam sistem [14].

Moses Decoder

Moses Decoder merupakan implementasi mesin penerjemah dengan pendekatan statistik (*statistical machine translation*) yang menggunakan konsep probabilitas dalam melakukan penerjemahan bahasa berdasarkan hasil analisis dari korpus paralel. Dapat dikatakan bahwa korpus paralel merupakan himpunan kalimat-kalimat dalam dua bahasa yang berbeda, yang berupa kalimat-kalimat selaras, yang mana pada setiap kalimat pada satu bahasa memiliki kesesuaian dengan kalimat yang diartikan pada bahasa lain [15].

Korpus paralel dapat digunakan oleh Moses Decoder setelah melakukan beberapa proses, berawal dari pra proses (*preprocessing*) berupa *Lowercase training data*, *Filter out long sentences*, dan *Tokenize training data* hingga kemudian *language modelling* (membangun model bahasa) dan terakhir *training* (pelatihan) pada sistem [14].

Penelitian ini menggunakan Moses Decoder untuk menghasilkan bahasa target yang memiliki nilai probabilitas tertinggi sebagai hasil dari penerjemahan bahasa sumber dengan menggunakan korpus paralel yang sudah diproses sebelumnya. Selain itu juga akan dibangun model bahasa menggunakan korpus monolingual bahasa terget dengan KenLM, lalu model translasi dan susunan kata untuk penerjemahan dibangun dengan GIZA++.

Korpus

Saat ini, konsepsi korpus telah berkembang dan memiliki cakupan makna yang luas. Sekarang, korpus juga bisa dikumpulkan secara elektronik. Untuk memahami konsep ini yang menjadi pertimbangan adalah pada tiga aspek. Pertama, korpus pada dasarnya ialah himpunan teks yang didapatkan secara elektronik, yang bisa untuk dianalisis secara otomatis maupun semi-otomatis. Aspek yang kedua bahwa korpus juga mencakup ujaran sehingga tidak hanya berisi kumpulan teks tertulis. Aspek yang ketiga, korpus juga dapat memuat dalam jumlah besar teks yang berasal dari berbagai sumber, contohnya dari beragam penulis dan pembicara yang berbeda pada berbagai subjek yang berbeda [16].

Korpus diartikan sebagai sekumpulan atau koleksi contoh teks tertulis atau lisan dengan bentuk yang bisa untuk dibaca dengan menggunakan seperangkat mesin dan bisa dianalisis menggunakan bentuk lain dari informasi linguistik bahasa yang berbeda [17].

Bentuk korpus yang dipakai dalam penelitian ini ialah korpus teks *chatting* bahasa sambas yang sudah dilakukan *tagging* perbaikan kata dan korpus paralel dokumen bahasa Melayu Sambas dan dokumen bahasa Indonesia. *Parallel corpus* atau korpus paralel sendiri merupakan pasangan korpus dengan isi kalimat-kalimat pada suatu bahasa dan terjemahannya. Korpus paralel akan digunakan untuk membangun mesin penerjemah statistik Moses Decoder, sedangkan korpus *tagging* digunakan untuk membangun mesin pra proses HMM.

K-Fold Cross Validation

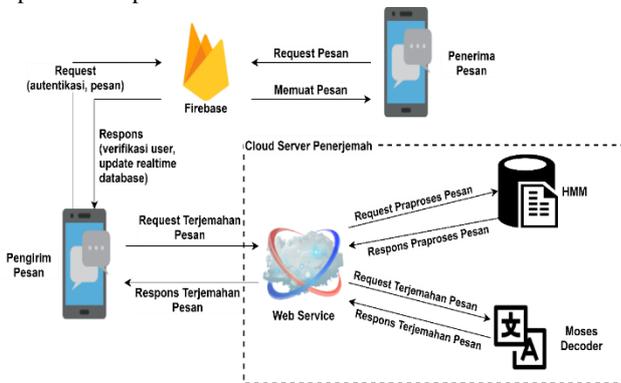
CV atau *cross validation* merupakan metode statistik yang bisa dipakai dalam menilai hasil kinerja suatu algoritma atau model dengan pembagian data ke dalam dua subset yaitu data *training*/pelatihan dan data uji/ validasi/evaluasi. Algoritma atau model dilatih menggunakan subset pelatihan dan dilakukan validasi menggunakan subset validasi. Tipe CV bisa dipilih berdasarkan dengan ukuran kumpulan data atau dataset. *Cross validation k-fold* biasanya dipakai karena mampu untuk mengurangi waktu perhitungan/komputasi dengan tetap mempertahankan ketepatan estimasi. Contoh untuk skema *10 fold cross validation* ditunjukkan pada Gambar 2.

hal ini dasar-dasar dalam penerjemahan bahasa. Analisis permasalahan ini akan dilakukan dengan mengidentifikasi karakteristik aplikasi *chatting* (aplikasi WhatsApp), kemudian mengidentifikasi karakteristik bahasa Melayu Sambas (penggunaannya pada aktivitas *chatting*).

Perancangan Sistem

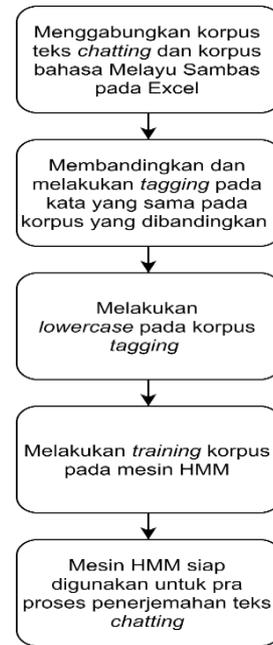
Perancangan sistem merupakan tahapan dalam menggambarkan kebutuhan sistem dan aktivitas yang terjadi dalam sistem. Pada perancangan sistem dilakukan perancangan arsitektur sistem, pembuatan diagram UML (*Unified Modelling Language*), dan perancangan antarmuka dari aplikasi *chatting*.

Arsitektur sistem ialah ilustrasi atau gambaran secara garis besar tentang bagaimana sistem dibangun lewat model-model yang berkorelasi antara satu dengan yang lain. Gambaran dari arsitektur sistem aplikasi *chatting* penerjemah yang dibangun diperlihatkan pada Gambar 4 berikut.



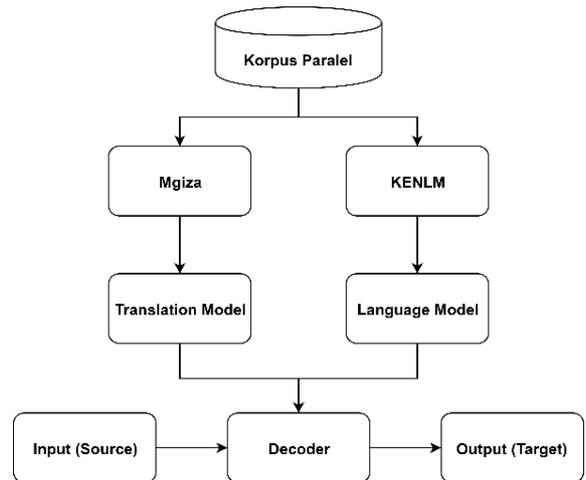
Gambar 4. Arsitektur Sistem Aplikasi *Chatting* Penerjemah

Pra proses oleh mesin HMM diperlukan untuk memperbaiki bahasa teks *chatting* yang memiliki ragam variasi abreviasi, slang ataupun pemanjangan kata sehingga menjadi bahasa Melayu Sambas yang lebih baik sebelum diterjemahkan ke bahasa Indonesia oleh Moses Decoder. Alur pembuatan mesin HMM yang akan digunakan untuk pra proses penerjemahan bahasa diperlihatkan pada Gambar 5 berikut.



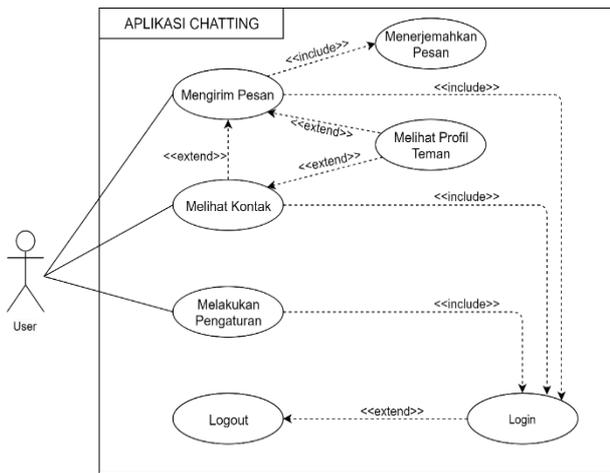
Gambar 5. Alur Pembuatan Mesin HMM

Arsitektur sistem mesin penerjemah statistik yang dibangun diperlihatkan pada Gambar 6 berikut.



Gambar 6. Arsitektur Mesin Penerjemah Statistik

Pada perancangan aplikasi *chat* penerjemah, jenis diagram UML yang digunakan berupa *Class diagram*, *Sequence diagram*, *Statechart diagram*, dan *Use Case diagram*. Diagram *use case* sendiri ialah pemodelan untuk perilaku aktor di dalam sistem aplikasi yang dikerjakan. Diagram *use case* menggambarkan hubungan atau interaksi antara aktor terhadap sebuah sistem. Gambar 7 berikut memperlihatkan *use case diagram* dari aplikasi yang dihasilkan.



Gambar 7. Use Case Diagram Aplikasi Chatting Penerjemah

Implementasi

Pada tahap implementasi dilakukan proses pembuatan server penerjemah dan proses pemrograman (*coding*). Pada pembuatan server penerjemah dilakukan instalasi dan konfigurasi mesin HMM dan Moses Decoder. Proses *coding* menggunakan bahasa pemrograman PHP untuk membuat *web service* server penerjemah dan bahasa pemrograman Java untuk membuat aplikasi *chatting* berbasis android.

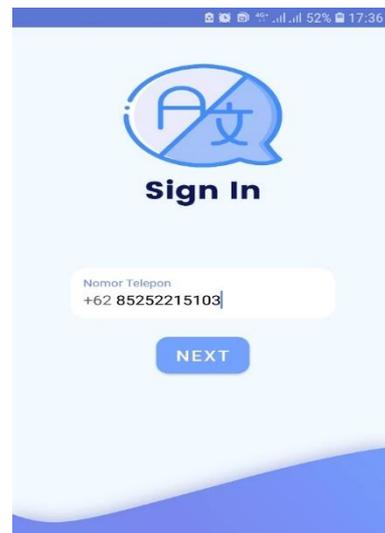
Pengujian Sistem

Pada pengujian sistem dilakukan tiga pengujian yaitu untuk menguji nilai akurasi hasil pra proses HMM, nilai akurasi hasil penerjemahan, dan keberhasilan pembuatan aplikasi. Pengujian nilai akurasi hasil pra proses HMM dilakukan dengan mencari nilai *accuracy*, *precision*, *recall*, dan *f-measure* dari hasil *tagging* mesin HMM dengan *trigram*. Pengujian nilai akurasi hasil penerjemahan dilakukan dengan pengujian hasil penerjemahan mesin translasi oleh BLEU dengan metode *k-fold cross validation* dan pengujian hasil penerjemahan mesin translasi oleh ahli bahasa. Sedangkan untuk menguji keberhasilan pembuatan aplikasi dilakukan dengan pengujian *black box*.

HASIL IMPLEMENTASI DAN PENGUJIAN

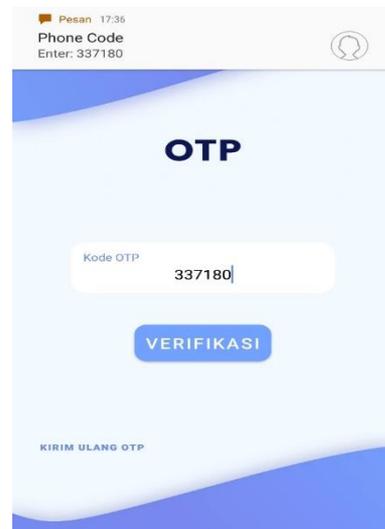
Hasil Implementasi

Hasil implementasi dari perancangan yang dilakukan adalah dalam bentuk aplikasi *chatting* penerjemah berbasis android. Pembangunan aplikasi dilakukan dengan bahasa pemrograman java serta menggunakan IDE (Integrated Development Enviroment) Android Studio. Server yang digunakan adalah Alibaba ECS (Elastic Cloud Server) yang di dalamnya di pasang *operating system* Ubuntu Server 16.04 LTS. Sedangkan untuk *service* penerjemah ditulis dengan bahasa pemrograman PHP. Berikut adalah beberapa tampilan yang dihasilkan dari perancangan antarmuka atau desain aplikasi, yang ditunjukkan oleh Gambar 8 sampai Gambar 15.



Gambar 8. Halaman Verifikasi Nomor Telepon

Halaman verifikasi nomor telepon merupakan rangkaian dari proses *login* pada aplikasi di mana *user* harus memasukkan nomor telepon aktif/valid yang digunakan pada perangkatnya.



Gambar 9. Halaman Verifikasi OTP

Setelah verifikasi nomor telepon berhasil maka Firebase akan mengirimkan sms berupa kode OTP, kemudian *user* dapat melakukan verifikasi untuk menyelesaikan proses *login*.



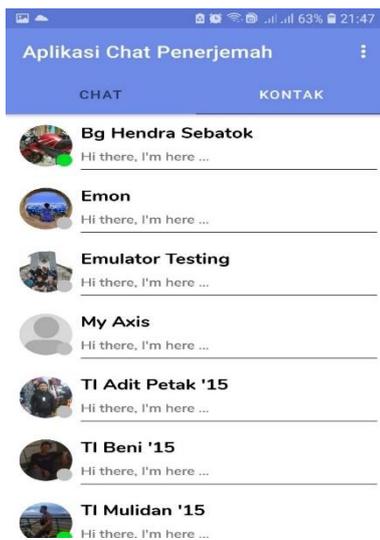
Gambar 10. Halaman Update Profil

Halaman *update* profil dapat diakses dengan dua cara yaitu setelah proses *login* berhasil dilakukan dan juga melalui halaman profil saya. Pada halaman *update* profil *user* dapat mengganti foto profil, nama, status, dan bahasa yang akan digunakan *user* dalam mengirim pesan. Jika *user* memilih bahasa Indonesia maka pesan tidak akan diterjemahkan, melainkan akan langsung dikirim ke penerima pesan. Sedangkan jika *user* memilih bahasa Sambas maka pesan yang dikirim *user* akan diterjemahkan ke bahasa Indonesia baru kemudian dikirim ke penerima pesan.



Gambar 11. Halaman Daftar Obrolan

Halaman daftar obrolan menampilkan daftar teman atau *user* lain yang melakukan obrolan dengan *user*. Informasi yang ditampilkan antara lain nama, foto profil, status aktif (apakah *user* lain sedang menggunakan aplikasi atau tidak), pesan terakhir yang dikirim, dan tanggal pesan terakhir dikirim. Teman atau *user* lain yang sedang aktif dapat dilihat dari indikator berupa lingkaran kecil berwarna hijau muda pada pojok kanan bawah dari foto profil teman, sedangkan untuk teman yang tidak aktif maka indikator akan menunjukkan warna abu-abu.



Gambar 12. Halaman Daftar Kontak

Halaman daftar kontak menampilkan daftar teman atau *user* lain yang menggunakan aplikasi *chat* penerjemah. Daftar kontak yang menampilkan daftar teman pada kontak perangkat *user* yang juga menggunakan atau terdaftar pada aplikasi *chat* penerjemah.



Gambar 13. Halaman Ruang Obrolan

Halaman ruang obrolan menampilkan percakapan atau obrolan antara *user* dengan *user* lain. Pada *bubble chat* yang dikirim *user* yang menggunakan bahasa Melayu Sambas, isi *bubble chat* dibagi menjadi tiga bagian yaitu pesan asli yang berada pada bagian atas, pesan hasil pra proses HMM pada bagian tengah, dan pesan yang sudah diterjemahkan oleh mesin penerjemah yang berada pada bagian bawah. Sedangkan *bubble chat* yang dikirim *user* yang menggunakan bahasa Indonesia hanya berisi pesan asli yang ditulis oleh *user* tersebut.



Gambar 14. Halaman Profil Saya

Halaman profil saya merupakan halaman yang menampilkan foto profil *user*, nama, status, dan bahasa yang digunakan. Pada halaman profil saya juga terdapat sebuah tombol yang berfungsi untuk mengarahkan *user* ke halaman *update* profil.

Gambar 15 atau gambar terakhir pada hasil implementasi adalah halaman profil teman, yaitu halaman yang menampilkan data profil teman (*user* lain). Halaman profil teman menampilkan foto profil, nama, nomor telepon dan status teman (*user* lain). Pada halaman profil teman terdapat sebuah tombol yang berfungsi untuk mengarahkan *user* ke halaman ruang obrolan sehingga *user* dapat mengirim pesan ke teman yang dipilih.



Gambar 15. Halaman Profil Teman

Hasil Pengujian

Pengujian Accuracy, Precision, Recall dan F-measure Pra Proses HMM

Pengujian *accuracy*, *precision*, *recall*, dan *f-measure* dilakukan untuk mengetahui nilai atau skor akurasi keberhasilan ketika melakukan *tagging* perbaikan atau pra proses bahasa Melayu Sambas *chatting* menjadi bahasa Melayu Sambas baku. Pengujian dikerjakan dengan membandingkan keluaran atau hasil *tagging* otomatis teks *chatting* oleh mesin HMM dengan hasil *tagging* manual teks *chatting*. Pengujian *accuracy*, *precision*, *recall*, serta *f-measure* dalam penelitian ini dilakukan menggunakan metode *k-fold cross validation* yaitu dengan skema *10-fold cross validation*. Pengujian dilakukan pada *10-fold* secara bergantian menggunakan bantuan *NLTK tools*, kemudian dihitung nilai rata-rata *10 fold* sebagai nilai akhir.

Mesin HMM dibangun dengan dua skenario korpus *tagging*, hasil dari uji *accuracy*, *precision*, *recall*, dan *f-measure* dari pra proses HMM dengan korpus *tagging* skenario pertama ditunjukkan oleh Tabel 1 di bawah.

Tabel 1. Hasil Pengujian Pra Proses HMM Skenario Pertama

Fold	Accuracy	Precision	Recall	F-Measure
1	0,646	0,840	0,839	0,840
2	0,913	0,814	0,794	0,804
3	0,912	0,836	0,801	0,818
4	0,939	0,850	0,839	0,844
5	0,951	0,875	0,872	0,874
6	0,910	0,766	0,775	0,771
7	0,927	0,826	0,834	0,830
8	0,928	0,827	0,834	0,831
9	0,923	0,845	0,832	0,838
10	0,948	0,864	0,868	0,866
Rata-Rata	0,9	0,834	0,829	0,832

Hasil pengujian dari pra proses HMM dengan korpus *tagging* skenario kedua memperoleh nilai *accuracy*, *precision*, *recall*, dan *f-measure* yang ditunjukkan oleh Tabel 2 berikut.

Tabel 2. Hasil Pengujian Pra Proses HMM Skenario Kedua

Fold	Accuracy	Precision	Recall	F-Measure
1	0,763	0,712	0,675	0,693
2	0,747	0,696	0,646	0,670
3	0,704	0,689	0,636	0,662
4	0,779	0,737	0,694	0,715
5	0,784	0,740	0,711	0,725
6	0,740	0,621	0,600	0,610
7	0,743	0,653	0,623	0,637
8	0,733	0,653	0,618	0,635
9	0,717	0,689	0,637	0,662
10	0,724	0,660	0,623	0,641
Rata-Rata	0,743	0,685	0,646	0,665

Pengujian Hasil Penerjemahan Mesin Translasi

Dilakukan pengujian pada hasil terjemahan mesin translasi menggunakan metode penilaian otomatis dengan BLEU dan uji manual oleh ahli bahasa (penutur bahasa Melayu Sambas). Dengan menggunakan BLEU (*Bilingual Evaluation Understudy*), pengujian hasil dari penerjemahan mesin translasi dilakukan untuk memperoleh keluaran dalam bentuk nilai akurasi hasil terjemahan. Evaluasi BLEU dilakukan dengan metode *k-fold cross validation* menggunakan skema *10-fold*.

Dibuat dua mesin translasi pada penelitian ini yaitu mesin penerjemah bahasa *chatting* Melayu Sambas ke bahasa Indonesia, serta mesin penerjemah bahasa Melayu Sambas (yang sudah diperbaiki) ke bahasa Indonesia. Hasil pengujian BLEU mesin pertama, mesin penerjemah bahasa *chatting* Melayu Sambas ke bahasa Indonesia (mesin pertama) ditunjukkan dengan Tabel 3 berikut.

Tabel 3. Hasil Pengujian BLEU Penerjemahan Bahasa Melayu Sambas (*chatting*) ke Bahasa Indonesia

Fold	Data Uji	Hasil Pengujian
1	1 – 450	34,24
2	451 – 900	46,97
3	901 – 1.350	53,2
4	1.351 – 1.800	57,85
5	1.801 – 2.250	51,09
6	2.251 – 2.700	46,56
7	2.701 – 3.150	45,27
8	3.151 – 3.600	45,57
9	3.601 – 4.050	36,9
10	4.051 – 4.500	39,55
Total	4500	457,2
Rata-Rata		45,72

Hasil dari pengujian BLEU dengan *10-fold cross validation* untuk mesin kedua, mesin penerjemah bahasa Melayu Sambas (yang sudah diperbaiki) ke bahasa Indonesia diperlihatkan oleh Tabel 4 berikut.

Tabel 4. Hasil Pengujian BLEU Penerjemahan Bahasa Melayu Sambas (Sudah Diperbaiki) ke Bahasa Indonesia

Fold	Data Uji	Hasil Pengujian
1	1 – 450	40,22
2	451 – 900	57,33
3	901 – 1.350	63,82
4	1.351 – 1.800	60,87
5	1.801 – 2.250	56,97
6	2.251 – 2.700	54,55
7	2.701 – 3.150	52,56
8	3.151 – 3.600	52,76
9	3.601 – 4.050	45,78
10	4.051 – 4.500	46,5
Total	4500	531,4
Rata-Rata		53,14

Berdasarkan hasil pengujian pada dua mesin penerjemah yakni mesin penerjemah langsung bahasa Melayu Sambas (*chatting*) ke bahasa Indonesia serta mesin penerjemah bahasa Melayu Sambas (yang sudah diperbaiki) ke bahasa Indonesia, menunjukkan bahwa mesin dengan penerjemahan bahasa Melayu Sambas yang sudah diperbaiki memiliki nilai akurasi yang lebih tinggi daripada penerjemahan langsung bahasa *chatting* Melayu Sambas ke bahasa Indonesia.

Pengujian manual diambil berdasarkan hasil uji otomatis dengan metode *k-fold cross validation* di mana dari hasil pengujian sebanyak 10 kali diambil nilai dengan hasil tertinggi. Adapun nilai tertinggi didapat pada *fold* ketiga mesin penerjemah bahasa Melayu Sambas (yang sudah diperbaiki) ke bahasa Indonesia, yaitu pada korpus uji kalimat 901 sampai kalimat 1.350 dengan nilai akurasi 63,82%. Pengujian lanjutan dilakukan pada *fold* dengan nilai akurasi tertinggi menggunakan BLEU dan skema *10-fold cross validation*. Pada *fold* yang memiliki nilai akurasi tertinggi dari pengujian lanjutan maka dilakukan pengujian manual yang melibatkan ahli bahasa. Hasil dari uji manual yang didapat diperlihatkan oleh Tabel 5 di bawah.

Tabel 5. Hasil Pengujian Terjemahan Mesin Penerjemah Oleh Ahli Bahasa

No.	Nama Ahli Bahasa	C	R	$P = \frac{C}{R} \times 100\%$
1.	Dendodi, S.Pd	304	350	86,86%
2.	M. Sutrisno, ST	307	350	87,71%

Pengujian Black Box

Pengujian *black box* ialah pengujian yang dapat dilakukan dengan pengamatan hasil dari eksekusi beberapa data pengujian tertentu serta pengujian fungsi-fungsi yang terdapat dalam perangkat lunak (*software*). Pengujian dengan *black box* dipilih karena tidak memerlukan akses ke kode program atau aplikasi dan dapat dilakukan ketika pengembangan telah selesai.

Pengujian *black box* perlu dilakukan untuk melihat respon yang diberikan oleh sistem ketika melakukan *entry* data dan apakah hasil eksekusi berhasil sesuai dengan yang diharapkan ataupun tidak berhasil. Pengujian yang dilakukan antara lain pengujian terhadap proses *login*, pengujian *update* profil, dan pengujian mengirim pesan.

Proses mengirim pesan terjadi pada halaman ruang obrolan, dan terdiri dari dua kondisi yaitu kirim pesan oleh *user* berbahasa Indonesia dan kirim pesan oleh *user* berbahasa Melayu Sambas. Tabel 6 berikut menunjukkan hasil pengujian kirim pesan.

Tabel 6. Hasil Pengujian Pengiriman Pesan

Fungsi	Kasus Uji	Hasil Eksekusi	Keterangan
Mengirim Pesan (Bahasa Indonesia)	Mengirim pesan kosong	Tidak berhasil	Pesan tidak terkirim
	Mengirim pesan “Saya sedang makan di rumah”	Berhasil	Pesan “Saya sedang makan di rumah” terkirim
Mengirim Pesan (Bahasa Melayu Sambas)	Mengirim pesan kosong	Tidak berhasil	Pesan tidak terkirim
	Mengirim pesan “Saye gek mkan drumah”	Berhasil	Pesan diterjemahkan dan dikirim: “Saye gek mkan drumah” (pesan asli) “saye agek makan dirumah” (hasil pra proses) “saya lagi makan di rumah” (terjemahan pesan)
	Mengirim pesan “Mmng tdok diluar tok, ngade2 juag tdok di dalm”	Berhasil	Pesan diterjemahkan dan dikirim: “Mmng tdok diluar tok, ngade2 juag tdok di dalm” (pesan asli) “memang tidok diluar tok, ngade- ngade juag tidok di dalam” (hasil pra proses) “memang tidur diluar nih, sembarangan juga tidur di dalam” (terjemahan pesan)

KESIMPULAN

Berdasar pada analisis dan hasil pengujian yang dilakukan terhadap mesin penerjemah statistik dan aplikasi yang dibangun, maka dapat ditarik kesimpulan bahwa pembuatan aplikasi *chatting* berbasis android yang mengimplementasikan mesin penerjemah statistik Moses Decoder dan dapat digunakan sebagai media komunikasi teks sudah berhasil dilakukan, di mana aplikasi *chatting* yang dibuat mampu menerjemahkan pesan bahasa Melayu Sambas ke bahasa Indonesia. Nilai akurasi penerjemahan langsung mesin penerjemah statistik bahasa Melayu Sambas (*chatting*) ke bahasa Indonesia ialah sebesar 45,72%, dan hasil akurasi penerjemahan dengan perbaikan kata mesin penerjemah statistik bahasa Melayu Sambas (baku) ke bahasa Indonesia ialah sebesar 54,14%. Sedangkan berdasarkan pengujian ahli bahasa untuk mesin kedua (penerjemahan dengan pra proses) memperoleh nilai persentase akurasi sebesar 86,86% dan 87,71%. Hasil pra proses HMM tertinggi (skenario pertama) memperoleh

rincian hasil sebesar 0,9 untuk nilai *accuracy*, 0,834 untuk nilai *precision*, 0,829 untuk nilai *recall*, serta 0,832 untuk nilai *f-measure*.

DAFTAR PUSTAKA

- [1] R. C. Cenderamata and A. N. Sofyan, "Abreviasi dalam Percakapan Sehari-Hari di Media Sosial: Suatu Kajian Morfologi," *Metahumaniora*, vol. 4, no. 1, 2019.
- [2] H. Sahara, "Implementasi Pengamanan Pesan Chatting menggunakan Metode Vigenere Cipher dan Cipher Block Chaining," vol. 3, no. 2, 2018.
- [3] F. Hardianto and B. Handaga, "APLIKASI GRUPCHAT DI ANDROID MENGGUNAKAN WEBSOCKET," 2015.
- [4] M. Syaokani, "Sistem Penerjemah Inggris-Indonesia Pada Aplikasi Chatting Berbasis Web Menggunakan Pendekatan Aturan," *TELKOMNIKA (Telecommunication Comput. Electron. Control.)*, vol. 8, no. 2, 2010.
- [5] N. F. Pratiwi, H. Sujaini, and R. D. Nyoto, "Pengembangan Antarmuka Mesin Penerjemah Statistik Multibahasa Berbasis Web," *J. Sist. dan Teknol. Inf.*, vol. 5, no. 1, 2017.
- [6] B. Hariyanto, "ISTILAH-ISTILAH KHUSUS DALAM CHATTING (Sebuah Analisis Sosiopragmatik)," *Adab. J. Bhs. dan Sastra*, vol. 9, no. 3, 2010.
- [7] D. Utami, "Karakteristik penggunaan bahasa pada status Facebook," 2010.
- [8] S. Priansya, "Normalisasi Teks Media Sosial Menggunakan Word2vec, Levenshtein Distance, dan Jaro-Winkler Distance," *Inst. Teknol. Sepuluh Nop.*, p. Tugas Akhir-KS141501, 2017.
- [9] I. G. B. B. Nugraha and R. D. Rizqullah, "Normalisasi Kata Tidak Baku yang Tidak Disingkat dengan Jarak Perubahan," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 3, p. 218, 2019.
- [10] M. E. B. Prasetyo, "Teori Dasar Hidden Markov Model," *Makal. Sekol. Tek. Elektro dan Inform. ITB*, 2010.
- [11] E. Loper and S. Bird, "NLTK: The Natural Language Toolkit," *Proc. COLING/ACL Interact. Present. Sess.* -, pp. 69–72, May 2002.
- [12] H. Tanuwijaya and H. M. Manurung, "PENERJEMAHAN DOKUMEN INGGRI-INDONESIA MENGGUNAKAN MESIN PENERJEMAH STATISTIK DENGAN WORD REORDERING DAN PHRASE REORDERING," *J. Ilmu Komput. dan Inf.*, vol. 2, no. 1, 2009.
- [13] I. Hadi, "UJI AKURASI MESIN PENERJEMAH STATISTIK (MPS) BAHASA INDONESIA KE BAHASA MELAYU SAMBAS DAN MESIN PENERJEMAH STATISTIK (MPS) BAHASA MELAYU SAMBAS KE BAHASA INDONESIA Ibnu Hadi," *J. Sist. dan Teknol. Inf.*, vol. 2, no. 3, 2014.
- [14] A. Ginting and N. AZ, "PENERJEMAH DUA ARAH BAHASA INDONESIA KE BAHASA DAERAH (KARO) MENGGUNAKAN TEKNIK STATISTICAL MACHINE TRANSLATION (SMT) SEBAGAI FITUR PADA SITUS WEB UNTUK MENINGKATKAN WEB TRAFFIC," *Telemat. MKOM Univ. Budi Luhur*, vol. 4, no. 1, 2012.
- [15] T. Wahyudinata, H. Sujaini, and R. D. Nyoto, "IMPLEMENTASI MESIN PENERJEMAH STATISTIK BERBASIS ANDROID DENGAN MOSES DECODER," *J. Sist. dan Teknol. Inf.*, vol. 3, no. 1, 2016.
- [16] M. Baker, "Corpora in Translation Studies," *Target. Int. J. Transl. Stud.*, vol. 7, no. 2, pp. 223–243, Jan. 1995.
- [17] T. McEnery, R. Xiao, and Y. Tono, *Corpus-based language studies: An advanced resource book*. Routledge, 2006.
- [18] P. Antinasari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1718–1724, 2017.
- [19] A. Hidayat, H. Sujaini, and R. Dwinyoto, "APLIKASI PENERJEMAH DUA ARAH BAHASA INDONESIA – BAHASA MELAYU SAMBAS BERBASIS WEB DENGAN MENGGUNAKAN DECODER MOSES," *J. Sist. dan Teknol. Inf.*, vol. 3, no. 3, 2015.