

PENERAPAN METODE C4.5 DALAM TINGKAT KEMAMPUAN BAHASA INGGRIS MAHASISWA

¹Ria Dewi Hudayani Sugara, ²Agustian, ³Imma Mardiyanti Visiam, ⁴M. Farhan, ⁵Dudih Gustian

¹Program Studi Pendidikan Bahasa Inggris ^{2,3,4,5}Program Studi Sistem Informasi

¹STKIP Bina Mutiara, ^{2,3,4,5}Sekolah Tinggi Teknologi Nusa Putra

¹Jl. Pembangunan Jl. Selakaso Kulon, Pasirhalang, Kec. Sukaraja, Kota Sukabumi

^{2,3,4,5}Jl. Raya Cibolang Kaler No.21 Kab. Sukabumi

e-mail: ¹ria@nusaputra.ac.id, ²agustian@nusaputra.ac.id, ³imma.mardiyanti@nusaputra.ac.id,

⁴farhan@nusaputra.ac.id, ⁵dudih@nusaputra.ac.id

Korespondensi: ¹ria.dewi@gmail.com

ABSTRAK

Dalam 5 tahun terakhir menunjukkan kemampuan bahasa Inggris bahasa Indonesia dari beberapa kategori sendiri. Khususnya tingkat kemampuan bahasa Inggris dari para mahasiswa khususnya semester akhir sangat rendah, bahkan bisa dikatakan sangat memprihatinkan. Oleh karena itu diperlukan penelitian terkait kemampuan bahasa Inggris di kalangan mahasiswa. Itulah sebabnya tingkat kemampuan bahasa Indonesia masih rendah, hal ini bisa dipahami karena sistem pendidikan di lingkungan kampus khususnya kurang mendukung hal itu dengan ditandai bahwa kedua tidak dianggap ilmu pengetahuan melainkan sebagai media komunikasi. Data diambil dari pengujian menggunakan tes TOEIC (Test Of English For International Communication) yang dikeluarkan oleh ETS (Educational Testing Service). Dengan menggunakan metode C4.5 akan lebih mudah untuk membuat klasifikasi tingkat kemampuan bahasa Inggris di kampus XYZ berdasarkan faktor yang diteliti. Penelitian ini memberi manfaat untuk melihat apa kekuatan tingkat kemampuan bahasa Inggris di kampus XY. Selain itu hasil yang diperoleh dapat membantu manajemen khususnya dosen dalam meningkatkan kinerja pembelajaran, sehingga tingkat kemampuan bahasa Inggris meningkat.

Kata kunci : TOEIC, ETS, metode C4.5

ABSTRACT

In the last 5 years, english proficiency Indonesian from several categories of its own. Especially the level of English ability of the students, especially the final semester is very low, bahkan can be said to be very concerning. Therefore, research is needed related to English language skills among students. That is why the level of ability Indonesian is still low, this can be understood because the education system in the campus environment in particular is less supportive of it by characterizing that the second is not considered science but rather as a medium of communication. The data was taken from tests using the TOEIC (Test Of English For International Communication) test issued by ets (educational testing service). Using the C4.5 method it will be easier to make a classification of english proficiency levels in XYZ village based on the factors studied. This study provides the benefit of seeing what the strength of english proficiency levels is on the XY campus. In addition, the results obtained can help management, especially lecturers in improving learning performance, so that the level of English language skills increases.

Keywords: TOEIC, ETS, C4.5 method

I. INTRODUCTION

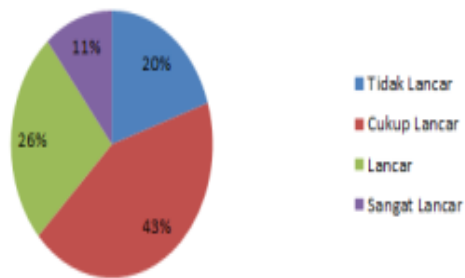
Dalam era globalisasi bahasa Inggris telah terjadi sejak dahulu dalam berbagai iptek,

ekonomi, politik, dan sosial budaya. Oleh karena itu, generasi penerus bangsa khususnya mahasiswa harus memiliki kemampuan bahasa Inggris yang

baik agar mampu bersosialisasi dengan bangsa lain di dunia sehingga Indonesia tidak terkucilkan dari pergaulan dunia internasional.

Namun terjadi permasalahan yakni masih rendahnya kemampuan bahasa Inggris dikalangan mahasiswa. Hal ini diperkuat menurut data dari Lembaga pendidikan dunia English First tentang indeks kemampuan berbahasa Inggris atau EF English Proficiency Index (EF EPI) di 44 negara. Bahasa Inggris di negara-negara itu bukan merupakan bahasa ibu atau pertama yang digunakan. Kemampuan bahasa Inggris di Indonesia berada sangat rendah di urutan ke-34, sedangkan Malaysia tembus di urutan ke-9 [1].

Tingkat kelancaran berbahasa Inggris di kampus XYZ

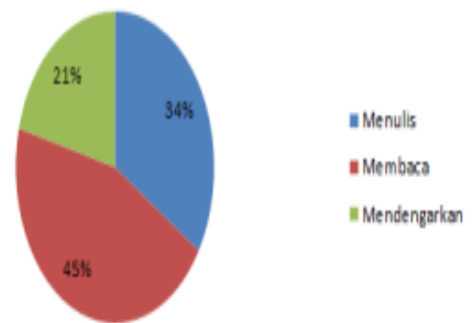


Gambar 1. Tingkat kelancaran berbahasa Inggris di kampus XYZ

Dalam gambar 1 diatas nampak bahwa tingkat kelancaran para mahasiswa kampus XYZ secara keseluruhan cukup tidak baik. Hal ini didiperlihatkan hanya 11% yang memiliki kemampuan sangat lancar dan terbanyak sekitar 43% cukup lancar. Ini tentunya menjadi masalah bagi pihak kampus XYZ untuk mencari solusi permasalahan tersebut, karena kemampuan berbahasa Inggris merupakan hal yang wajib dimiliki oleh setiap mahasiswa.

Jika dikelompokkan menurut tingkat pemahaman sesuai dengan standar ETS, maka terdapat tiga kelompok besar yang wajib dikuasai oleh mahasiswa. Sekitar 45% mahasiswa sebagian besar dapat menguasai dalam membaca, namun 34% memiliki kemampuan menulis dan hanya 21% dalam mendengarkan.

Tingkat Pemahaman Bahasa Inggris di Kampus XYZ



Gambar 2. Tingkat pemahaman bahasa Inggris sesuai bidang penguasaan standar ETS

Penelitian ini menggunakan metode C4.5 guna mengelompokkan kemampuan bahasa Inggris mahasiswa sesuai dengan level dasar, menengah dan tinggi.

Diharapkan penelitian ini dapat memberi masukan kepada pihak kampus XYZ dalam mencari cara bagaimana meningkatkan kemampuan berbahasa Inggris mahasiswanya serta memberi solusi kenapa tingkat kemampuan bahasa Inggris mahasiswanya rendah.

II. TINJAUAN PUSTAKA

2.1 Penelitian Terkait

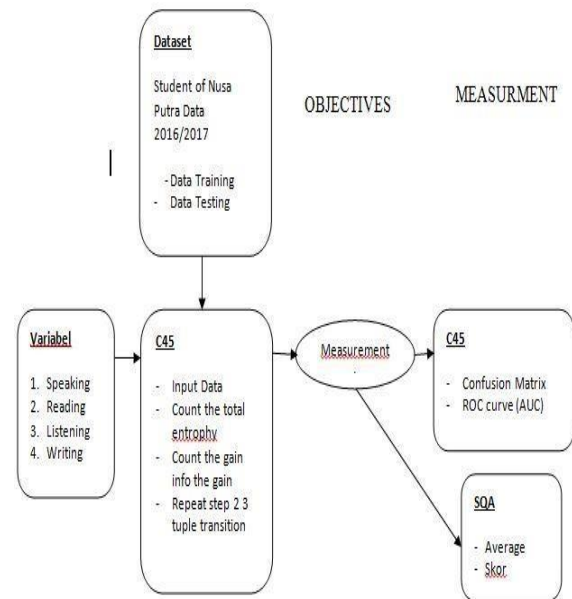
Thakaa Z. Mohammad dkk (2014). Dengan judul "Penelitian cerdas pendidikan data mining model klasifikasi untuk mengajar bahasa Inggris untuk siswa pelajar lambat". Peneliti menyajikan penelitian untuk membangun model klasifikasi klasifikasi data pendidikan cerdas yang dirancang untuk mengajar bahasa Inggris bagi siswa pelajar lambat belajar untuk mendapatkan yang terbaik dan mencapai peningkatan kinerja mereka. Usulan tersebut dicapai melalui penerapan empat tahap utama, diantara tahapannya yaitu: 1. pengembangan kursus bahasa Inggris, 2. desain & implementasi basis data, 3. mengumpulkan data dan fase, 4. menerapkan teknik. Metode pengajaran bahasa Inggris menggunakan e-course, dimana hasilnya menunjukkan bahwa siswa yang

mengikuti pola (membaca, menulis, berbicara dan tata bahasa) mencapai prestasi akademik tertinggi >79,5%. Populasi sampel dipilih secara acak dan bertujuan untuk mencapai 500 peserta yang menggunakan perangkat lunak e-learning, tetapi hanya 456 siswa yang tidak dapat menyelesaikan kursus dengan baik dan tuntas. Sistem ini secara signifikan membantu guru dalam belajar dan menghemat waktu [2].

Ina Sukaesih dan Junaedi (2012). Melakukan penelitian dengan judul "English in service industry indonesia measured using the common framework for Europe". Penelitian ini menggunakan kerangka kerja umum skala *kuantitatif deskriptif* untuk Eropa. Rata-rata pada kualitas bahasa dan kemampuan strategi didominasi oleh kemampuan penguasaan kosakata yaitu rata-rata 3,10 dan 3,25. Dimana penguasaan kosakata ini telah berhasil digunakan dalam kemampuan menulis untuk menyalakan rata-rata 3,15 dan menulis biodata pada dasarnya dengan nilai rata-rata 3,15. Menulis surat kepada editor menunjukkan hasil rendah yaitu 2,70. Kemampuan produk dengan rata-rata 2,96, dengan rata-rata terendah adalah kemampuan untuk melihat dan meringkas berita intisari, wawancara dan dokumenter 2,80. Kemampuan untuk mendengar rata-rata 3,20 dan kemampuan membaca rata-rata 3,17. Seluruh pekerja di Jabodetabek menunjukkan cukup banyak, yaitu 3,05. Rata-rata kemampuan yang umum dalam pendengaran, membaca, interaksi lisan, produksi lisan, strategi dan kualitas bahasa dan menulis [3].

Alfa Saleh (2015). Melakukan penelitian dengan judul "Penerapan Data Mining dengan Metode Klasifikasi Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa dalam Mengikuti English Proficiency Test (Studi Kasus : Universitas Potensi Utama)". Dari 50 data mahasiswa yang mengikuti English Proficiency Test telah diuji dengan metode Naïve Bayes, didapatkan hasil persentase sebesar 98% untuk keakuratan klasifikasi. Diketahui dari 50 data yang diuji terdapat 49 data yang berhasil diklasifikasikan dengan benar [4].

2.2 Kerangka berfikir



Gambar 3. Kerangka berfikir yang dibuat

Dalam gambar 3 diatas variable yang diteliti terdiri dari 4 variabel yaitu : speaking, reading, listening dan writing. Dimana ke 4 hal tersebut merupakan variabel yang termasuk dalam standar penilaian ETS. Dataset yang digunakan mahasiswa XYZ tahun akademik 2016/2017 yang terbagi menjadi 70% data training dan 20% data testing dan 10 % data validasi. Luaran yang dihasilkan merupakan akurasi C4.5 dengan pengukuran confusion metric dan ROC.

III. METODOLOGI PENELITIAN

3.1 Data Penelitian

Data yang diolah proses data berasal dari data primer yang diperoleh langsung dari siswa tahun akademik 2016/2017 dengan total 996 sampel.

3.2 Metode C4.5

Metode C4.5 menghitung akar pohon, untuk mencetak rumus yang digunakan entropi [7].

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i$$

Keterangan :

S= Himpunan kasus

n = jumlah partisi S

P_i= proporsi S_i terhadap S

Untuk skor mendapatkan perhitungan nilai keuntungan untuk menemukan entropi nilai dengan rumus sebagai berikut:

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i)$$

Keterangan :

S = Himpunan Kasus

A = Fitur

n = jumlah partisi atribut A

|S_i| = Proporsi S_i terhadap S

|S| = jumlah kasus dalam S

IV. HASIL DAN PEMBAHASAN

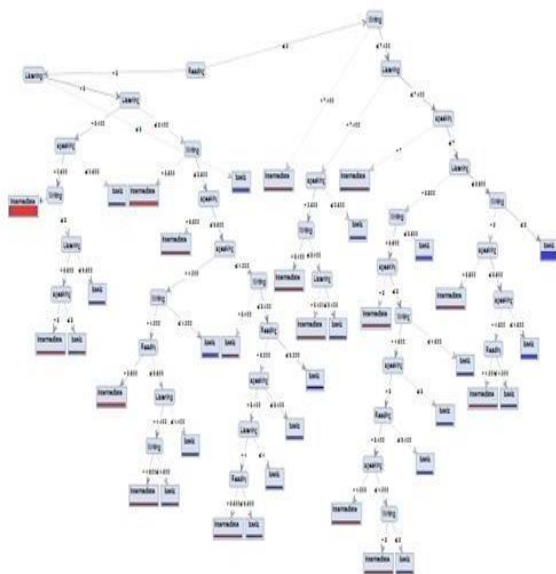
4.1 Analisis Data

Hasil analisis mahasiswa kampus XYZ di semester awal menggunakan nilai rata-rata masing-masing indikator. Memperoleh jumlah nilai di atas 5,00 sekitar 128 mahasiswa untuk indikator kemampuan berbicara, sedangkan indikator mendengarkan memperoleh jumlah oleh 214 siswa, untuk indikator membaca diperoleh jumlah sekitar 195 siswa dan jumlah penulisan indikator sekitar 190 siswa dari total 399 siswa yang diambil.

4.2 Analisis dengan metode C4.5

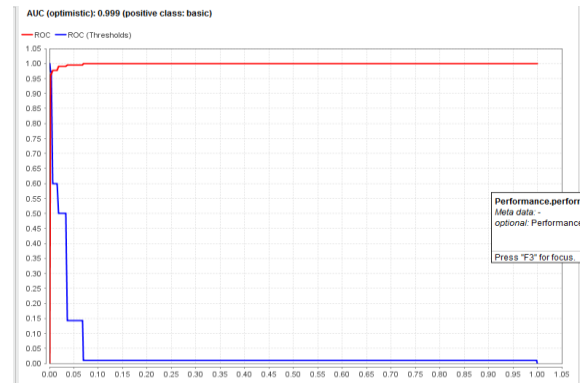
1. Data Training

Setelah mengatur sesuai dengan kelasnya masing-masing, jika mahasiswa digolongkan dengan indikasi dasar, menengah dan muka dicampur dengan penambang cepat pelatihan data.



Gambar 4. Pohon decision dari pelatihan data

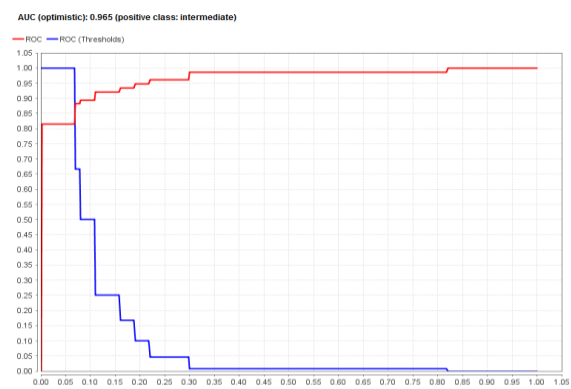
Pada gambar 4 pada jelas bahwa ada 33 aturan pohon dengan jumlah kelas dasar sekitar 14 aturan, intermediete sekitar 17 aturan dan maju sebanyak 2 aturan.



Gambar 5. ROC data training

2. Data Testing

Langkah selanjutnya adalah melakukan proses validasi, dimana langkah tersebut bertujuan untuk menguji aplikasi data mining yang telah kita buat yang memperoleh persentase sistem keakuran yang dibangun oleh . Sedangkan tingkat akurasi matriks confusionsion dalam menghasilkan presisi sekitar 89 persen sekitar 87,50% dan racall sekitar 85,00%. Adapun tingkat akurasi auc dalam menghasilkan validasi sektar 0,999 % ini berarti termasuk klasifikasi yang sangat baik.



Gambar 6. ROC Testing

4.3 Pembahasan

Seperti yang ditunjukkan pada hasil analisis,menjukkan bahwa pendidikan menempati urutan pertama dari perangkian dengan pohon keputusan, diikuti oleh pekerjaan pada setiap kepala keluarga dan pendapatan dan sebagainya. Hal ini dapat dipahami karena mungkin faktor pendidikan sangat mempengaruhi pada

gelar/pekerjaan, dan itu jelas membuat efek bagi masyarakat dalam kesejahteraan.

Selanjutnya setiap kepala rumah tangga dapat menghasilkan pendapatan patokan. Kemudian dalam klasifikasi segmen dengan metode C45 dilakukan proses pembentukan pohon model dari pelatihan data. Dimana tingkat akurasi data dari pengujian, dilanjutkan dengan melakukan proses validasi kedua set data ini. Kemudian pada klasifikasi segmen dengan metode C45, dilakukan proses pembentukan pohon model dari pelatihan data dengan tingkat akurasi data dari pengujian, Melakukan proses validasi kedua set data ini, dengan tingkat akurasi 98,02% data pengujian dan auc 0,999% dan nilai validasi auc sekitar 0,989%, akurasi 96,27%, presisi 97,05% dan ingat 98,16%.

V. PENUTUP

5.1 Kesimpulan

Implementasi metode C4.5 untuk melakukan prediksi tingkat bahasa siswa Inggris mahasiswa kampus XYZ telah berhasil melakukan hasil dari tingkat keberhasilan tingkat prediksi kemampuan mahasiswa kampus XYZ telah diukur dalam keputusan pohon adalah dicatat bahwa akurasi pengujian data adalah 97,98% indikasi menengah sekitar 98,29%, Indikasi dasar tentang 97,75 % senilai masing-masing persisi kelas 97,18 % untuk prediksi indikasi menengah dan 98,64 % untuk prediksi indikasi dasar. Untuk penelitian lebih lanjut, harus dilakukan uji coba data sampel menggunakan lebih akurat bahwa aplikasi dapat ditingkatkan.

5.2 Saran

Untuk mengoptimalkan tingkat akurasi hendaknya digunakan metode optimasi seperti algoritma Genetika, PSO dan Algoritma Semut dan sebagainya.

DAFTAR PUSTAKA

- [1] A. Mulyadi. "Kemampuan Bahasa Inggris di Indonesia Rendah", Wired, 26/09/2011., [Online].
<https://edukasi.kompas.com/read/2011/09/26/21320318/Kemampuan.Bahasa.Ingggris.di.Indonesia.Rendah> [Diakses: 25 Januari 2015.
- [2] T. Z. Mohammad. "An Intelligent Education Data Mining Classification Model for Teaching English For Slow Learner Students". 2014 Vol 2 (8).
- [3] I.Sukaesih. "Kemampuan Bahasa Inggris Dalam Industri Jasa Indonesia Diukur dengan Menggunakan Skala Pengukuran Kerangka Kerja Eropa". 2012.
- [4] A.Saleh. "Penerapan Data Mining Dengan Metode Klasifikasi Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa Dalam Mengikuti English Proficiency Test (Studi Kasus : Universitas Potensi Utama)", Conference: Konferensi Nasional Sistem Informasi, Universitas Klabat, Manado, Indonesia, Volume: Februari 2015.