



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB II

LANDASAN TEORI

2.1. *Education First English Proficiency Index (EF EPI)*

Education First English Proficiency Index (EF EPI) adalah sebuah penilaian terhadap tingkat kemampuan berbahasa Inggris sebuah negara. EF EPI juga digunakan untuk melihat seberapa baik kemampuan bahasa Inggris sebuah negara dibandingkan dengan negara lainnya. EF EPI ini didasarkan kepada hasil dari tes bahasa Inggris yang dilakukan oleh EF pada negara-negara setiap tahunnya. Hasil dari tes tersebut kemudian dijadikan data untuk menilai kemampuan bahasa Inggris suatu negara.

EF EPI pertama kali dipublikasikan pada tahun 2011 berdasarkan data dari 1.7 miliar hasil tes dan berlanjut setiap tahunnya hingga yang terbaru dan ke-8 yang dipublikasikan pada tahun 2018. Pada hasil EF EPI yang dirilis pada tahun 2018, digunakan data dari 1.3 miliar tes di tahun 2017. 85 negara masuk ke dalam EF EPI ke-8 ini. Syarat untuk sebuah negara masuk ke dalam EF EPI adalah memiliki hasil tes sebanyak 400 atau lebih. ("EF EPI 2018 - Indeks Kecakapan Bahasa Inggris EF", 2019)

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

2.2. TOEIC

TOEIC (*Test of English International Communication*) adalah tes bahasa Inggris yang dibuat dengan tujuan untuk menilai kemampuan berbahasa Inggris seseorang dalam dunia kerja secara Internasional. Tes TOEIC terdiri dari TOEIC *Listening & Reading Test* dan TOEIC *Speaking & Writing Test*. TOEIC *Listening & Reading Test* berlangsung selama 2 jam (45 menit untuk *listening* dan 75 menit untuk *reading*) dengan 200 pertanyaan pilihan ganda yang terbagi untuk kedua bagian tersebut. Skala nilai untuk bagian *listening* dan *reading* masing-masing adalah 5-495 sehingga skala nilai total untuk TOEIC *Listening & Reading Test* adalah 10-990. Sertifikat untuk TOEIC *Listening & Reading Test* terbagi ke dalam 5 warna, oranye (10-219), coklat (220-469), hijau (470-729), biru (730-859), emas (860-990).

TOEIC *Speaking & Writing Test* memberikan nilai terpisah dengan skala 0-200. Peserta tes dapat mengambil tes secara terpisah, *speaking* saja atau *writing* saja. Tes ini dibuat untuk menggambarkan penggunaan bahasa Inggris dalam dunia kerja secara umum sehingga peserta tes tidak perlu memiliki pengetahuan khusus tentang sebuah bidang pekerjaan. Bagian tes *speaking* berlangsung selama 20 menit sementara bagian *writing* berlangsung selama 60 menit. ("TOEIC", 2019)

2.3. Data Processing

Data processing adalah kegiatan mengumpulkan dan memanipulasi data. *Data processing* dapat juga dikatakan sebagai *information processing*. *Data processing* terbagi dalam beberapa jenis berdasarkan area aplikasi yaitu *scientific data processing* dan *commercial data processing*. *Scientific data processing* adalah

kegiatan *data processing* yang dilakukan pada bidang akademik dan riset. *Commercial data processing* adalah kegiatan *data processing* yang dilakukan menggunakan jumlah data yang banyak, biasanya dilakukan dengan pemrosesan secara sederhana, dan hasil dari *data processing* jumlah ini juga berjumlah banyak.

Data processing berdasarkan cara pemrosesan data dibagi ke dalam *manual data processing*, *mechanical data processing*, dan *electronic data processing*. *Manual data processing* adalah kegiatan *data processing* yang dilakukan secara manual tanpa bantuan alat. *Mechanical data processing* adalah kegiatan *data processing* yang dilakukan menggunakan alat-alat sederhana seperti kalkulator. *Electronic data processing* adalah kegiatan *data processing* modern yang dilakukan dengan bantuan komputer untuk mencapai hasil *data processing* yang cepat dan akurat. (Chung & Aring, 2018)

2.4. Data Cleansing

Data cleansing adalah proses mendeteksi dan membenarkan ketidaktepatan yang ada pada tabel atau database. Setelah proses *cleansing* selesai, kumpulan data seharusnya konsisten dengan kumpulan data lainnya yang bersifat mirip pada sistem yang sama. Ketidakkonsistenan pada data dapat disebabkan oleh karena kesalahan input oleh pengguna, kerusakan pada transmisi atau penyimpanan data, atau perbedaan definisi pada kumpulan data yang berbeda. Perbedaan antara *data cleansing* dan *data validation* adalah *data validation* dilakukan pada waktu data diinput sementara *data cleansing* dilakukan setelah data telah menjadi kumpulan data.

Suatu cara *data cleansing* yang sering dilakukan adalah *data enhancement* yaitu menambahkan informasi yang berkaitan dengan data tersebut.

Proses *data cleansing* yang baik akan menghasilkan data berkualitas tinggi. Data berkualitas tinggi harus mencukupi beberapa kriteria yaitu *validity*, *criteria*, *completeness*, *consistency*, dan *uniformity*. Proses *data cleansing* meliputi *data auditing*, *workflow specification*, *workflow execution*, *post-processing and controlling*, *parsing*, *data transformation*, *duplicate elimination*, dan *statistical methods*. (Hailu, 2015)

2.5. Data Transformation

Data transformation adalah proses mengubah data dari menjadi bentuk yang diperlukan. Proses *data transformation* data bersifat simpel atau kompleks berdasarkan dari data yang dibutuhkan. Proses-proses dalam kegiatan *data transformation* antara lain adalah *data discovery*, *data mapping*, *code generation*, *code execution*, dan *data review*. (Hailu, 2015)

2.6. Data Modeling

Data modeling adalah proses untuk menentukan dan menganalisa kebutuhan data untuk membantu dalam proses bisnis. Proses *data modeling* terdiri dari perekaman kebutuhan data sebagai *conceptual data model* yang digunakan untuk membicarakan kebutuhan mula-mula. Setelah itu *conceptual data model* diubah menjadi *logical data model* yang menyimpan struktur data yang dapat diimplementasi pada *database*. Terakhir *logical data model* diubah menjadi *physical*

data model yang memasukkan data ke dalam tabel. Menurut Jeffrey L. Whitten ada 2 jenis *data modeling* yaitu *strategic data modeling* dan *data modeling* pada *system analysis*. (Ribeiro, Silva & da Silva, 2015)

2.7. Data Visualization

Data visualization adalah kegiatan komunikasi informasi dari data melalui visual. Untuk mengkomunikasikan informasi dari data, kegiatan *data visualization* dilakukan menggunakan hal-hal seperti *statistical graphics*, *plots*, dan *information graphics*. Visualisasi yang baik akan membantu pengguna menganalisa data yang ada sehingga data menjadi lebih mudah dimengerti dan digunakan. Menurut Fernanda Viegas dan Martin M. Wattenberg, *data visualization* yang ideal seharusnya tidak hanya mengomunikasikan data secara jelas tetapi juga harus dapat menarik perhatian dari orang-orang yang melihat. (Wang, Zhao, Cao, Qu & Li, 2018)

2.8. SPSS Statistics

SPSS Statistics adalah software untuk proses analisa statistik yang awalnya dibuat oleh *SPSS Inc.* *SPSS Statistics* kemudian diubah menjadi *IBM SPSS Statistics* setelah diperoleh oleh IBM. Statistik-statistik yang ada dalam *SPSS* adalah statistik deskriptif, statistik *bivariate*, prediksi untuk hasil *numerical*, prediksi untuk mengidentifikasi grup, analisis dan simulasi geospasial, ekstensi R untuk Python. ("SPSS Software", 2019)

2.9. Korelasi

Korelasi adalah hubungan secara statistik antara suatu faktor dengan faktor lainnya. Korelasi mengindikasikan adanya keterkaitan antara suatu faktor dengan faktor lain dalam bentuk statistik, tetapi korelasi antara dua hal atau lebih tidak bisa diartikan sebagai adanya sebab akibat antara hal-hal tersebut. (Zheng et al., 2014)

2.10. Knowledge Data Discovery

Knowledge Data Discovery (KDD) adalah proses yang dilakukan untuk menemukan informasi dalam data dan menggunakan metode *data mining*. Proses KDD terdiri dari *Selection, Preprocessing, Transformation, Data Mining, Interpretation/Evaluation*. (Hailu, 2015)

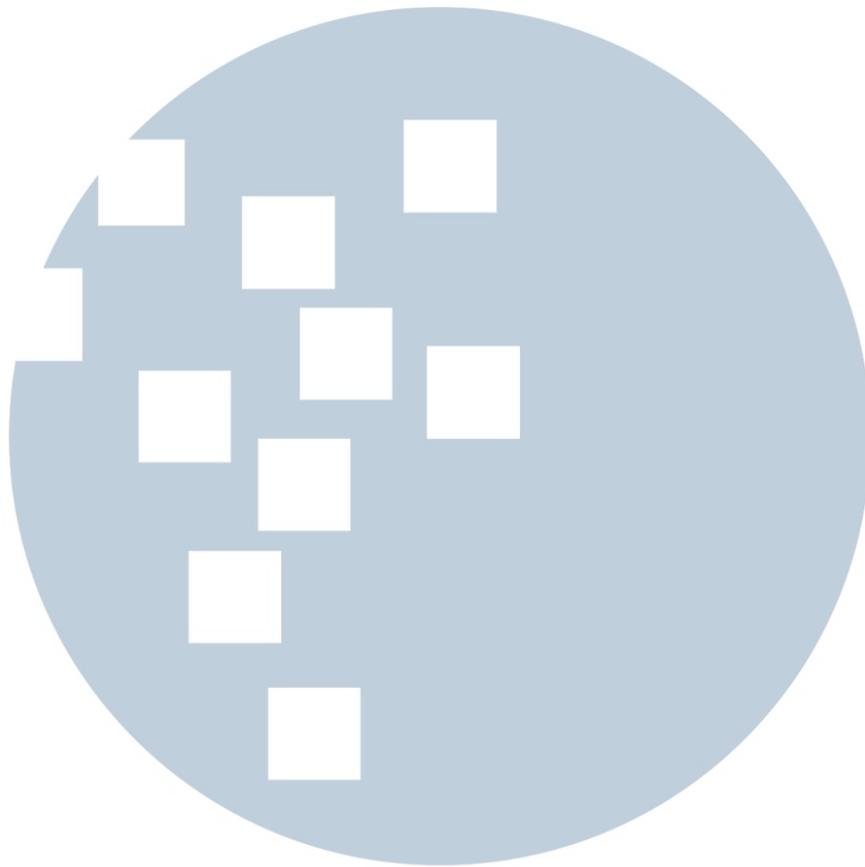
2.11. CRISP DM

CRISP DM (*cross-industry standard process data mining*) adalah model proses *data mining*. Pada tahun 2015 IBM mengeluarkan metodologi baru yaitu *Analysis Solutions Unified Method for Data Mining/Predictive Analysis* (ASUM DM) yang adalah pengembangan dari CRISP *data mining*. Kegiatan-kegiatan dalam proses CRISP DM adalah *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*. (Hailu, 2015)

2.12. Penelitian Pendahuluan

No	Judul	Pengarang	Kesimpulan
1.	ANOVA untuk Analisis Rata-rata Respon Mahasiswa Kelas <i>Listening</i> <i>Prosiding SNMPM</i> Universitas Sebelas Maret 2013 <i>Volume</i> 2	Noviataro Fury Pritasari, Hanna Arini Parhusip, Bambang Susanto	Meneliti apakah ada perbedaan respon pada kelas <i>listening</i> menggunakan ANOVA dengan <i>one-way repeated measures</i> dan <i>two-way repeated measures</i>
2.	<i>Association Between Physician Online Rating and Quality of Care</i> <i>Journal of Medical Research</i> 18	Kanu Okike, Taylor K Peter-Bibb, Kristal C Xie, Okike N Okike	Meneliti apakah ada hubungan antara <i>rating online</i> dokter dengan kualitas pelayanan dokter

UNIVERSITAS
MULTIMEDIA
NUSANTARA



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA