



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, 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 III METODE PENELITIAN

3.1. Objek Penelitian

Objek penelitian ini adalah diambil dari hasil nilai tes TOEIC mahasiswa Universitas Multimedia Nusantara (UMN). Data nilai tersebut diperoleh dari IT *Department* Universitas Multimedia Nusantara. Data yang diperoleh berisikan data nilai *reading score*, *listening score*, *total score*, ID, Program studi, asal sekolah, dan *address*. Data tersebut diambil berdasarkan hasil test TOEIC mahasiswa UMN dari bulan Januari 2019 sampai dengan bulan Mei 2019.

Data yang dimiliki departemen IT Universitas Multimedia Nusantara merupakan data resmi yang berisikan informasi mengenai peserta tes TOEIC. Data tes TOEIC dipilih karena data tersebut resmi hasil tes yang di berikan dari lembaga *Continuing Education Department* (CED) UMN, sehingga dipastikan data yang dimiliki oleh departemen IT tersebut valid.

Data yang diteliti adalah rata-rata nilai TOEIC mahasiswa umn, yang dibagi berdasarkan program studi (prodi) dan juga daerah asal. Prodi yang diteliti adalah prodi Sistem Informasi, Teknik Komputer, Akuntansi, Informatika, Manajemen, Desain Komunikasi Visual, Film, dan Ilmu Komunikasi. Proses pengumpulan data daerah akan dipilih berdasarkan data kota mahasiswa UMN. (logistic regression bandingin)

3.2. Metode Penelitian

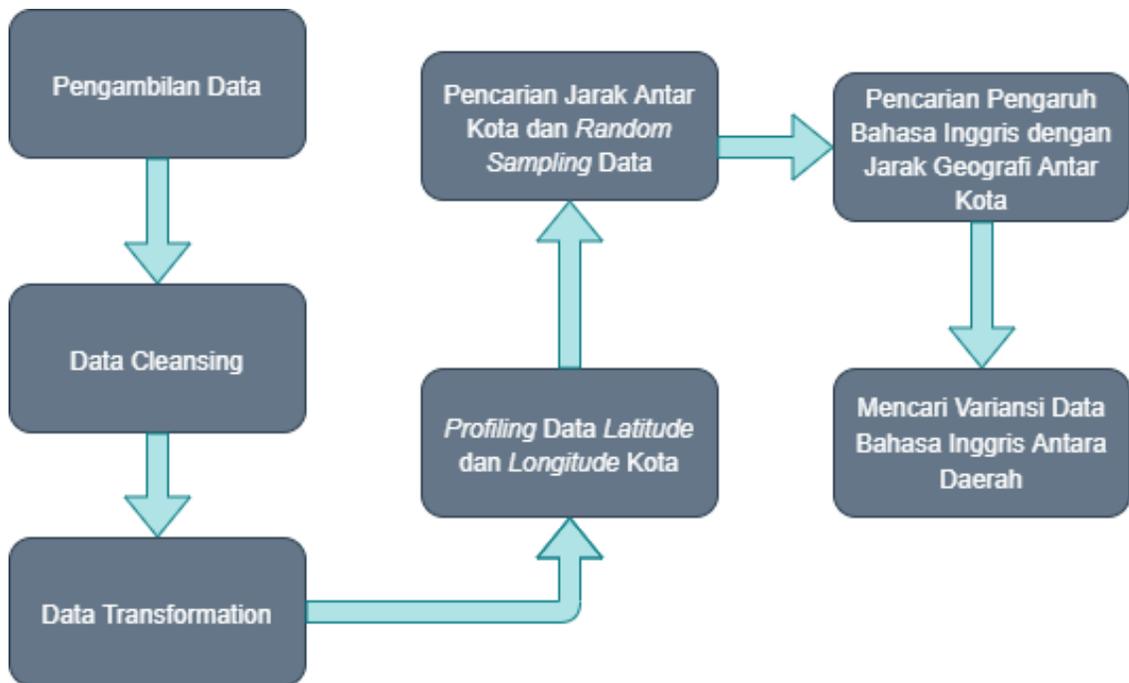
Tabel berikut ini merupakan perbandingan untuk melihat perbedaan diantara metode *linear regression* dan *logistic regression*.

Tabel 3.1 Tabel perbandingan metode

<i>Distinction</i>	<i>Algorithms</i>	
	<i>Linear Regression</i>	<i>Logistic Regression</i>
<i>Functionality</i>	<i>Predict continuous dependent variable</i>	<i>Predict categorical dependent variable</i>
<i>Solving type</i>	<i>Regression Problem</i>	<i>Classification Problem</i>
<i>Accuracy</i>	<i>Higher</i>	<i>Lower</i>

Penelitian ini dilakukan dengan menggunakan metode regresi linear, berdasarkan perbandingan yang dilihat dari tabel 3.1 metode regresi linear adalah metode yang lebih baik untuk menemukan kesinambungan atau keterkaitan diantara satu tempat dengan tempat lainnya. Penggunaan metode regresi linear juga disertakan dengan penelitian sebelumnya yang menggunakan metode regresi linear untuk mencari keterkaitan antara lagu daerah di Indonesia dengan melihat jarak antara satu daerah dengan ciri-ciri lagu daerah (Kuanca, 2019). metode regresi linear digunakan untuk mencari pola atau hubungan yang ada. Hubungan yang ingin ditemukan adalah keterkaitan daerah dengan kemampuan berbahasa Inggris. Pada hal ini akan mencari apakah sebuah daerah memiliki keterkaitan dalam menguasai kecapakan dalam berbahasa Inggris, melihat persaingan pasar bebas MEA yang sudah berjalan di Indonesia.

Tahapan yang akan dilakukan dalam melakukan penelitian ini terbagi menjadi 6 tahap seperti pada gambar 3.1 dibawah ini dengan pembahasan sebagai berikut: (1) pengambilan data, (2) *data cleansing*, (3) *data transformation*, (4) *profiling* data *latitude* dan *longitude* kota, (5) pencarian jarak antar kota dan *random sampling* data, (6) pencarian Pengaruh Penguasaan Bahasa Inggris dengan Jarak Geografi Antar Kota, dan (7) mencari variansi data bahasa Inggris antara daerah. Setelah itu hasil metode penelitian yang didapat akan dibahas setelah perhitungan model regresi dengan *Oriented Least Structure* atau OLS dan pencarian variansi data dengan Anova selesai.



Gambar 3.1 Flowchart alur pengumpulan data

3.2.1 Pengambilan Data

Data yang digunakan di dalam penelitian ini adalah data hasil tes TOEIC mahasiswa UMN, Data TOEIC ini didapat dari IT *Department* UMN. TOEIC merupakan sebuah tes yang diberikan UMN kepada mahasiswa melalui lembaga pengembangan pendidikan yang dinamakan *Continuing Education Department* atau dikenal dengan nama CED. Data yang diterima berbentuk *file excel* yang di dalamnya memiliki beberapa kolom tabel, isi dari tabel tersebut adalah ID, program studi, *last school attend id*, asal sekolah, *score listening*, *score reading*, *total score*, dan alamat rumah mahasiswa. Total data yang diterima berjumlah 5,142 baris, tetapi tidak semua data dipakai karena terdapat data yang *duplicate* dan data yang diteliti adalah data daerah Jabodetabek saja.

3.2.2 Data Cleansing

Proses selanjutnya adalah melakukan *data cleansing*, proses ini dilakukan di dalam *Micorsoft Excel*. Data yang akan mengalami proses penghapusan pada *data cleansing* adalah data *duplicate* berdasarkan data ID. Proses *cleansing* selanjutnya adalah menghapus kolom yang tidak terpakai selama penelitian, kolom yang dihilangkan adalah kolom asal sekolah dan *last school ID*. Pada proses *cleansing* juga akan melakukan koreksi terhadap data yang tidak lengkap, hal ini dilakukan untuk mendukung proses penelitian dengan memperoleh data yang berkualitas

3.2.3 Data Transformation

Berdasarkan data yang diperoleh, data alamat rumah mahasiswa yang didapat merupakan alamat lengkap yang terdiri dari nama jalan, blok rumah, nomor rumah, RT dan RW, kecamatan, kota, dan provinsi. Transformasi data dilakukan karena beberapa baris data memiliki urutan yang tidak teratur dan memiliki susunan alamat yang berbeda-beda, sehingga data yang digunakan harus ditransformasi agar lebih mudah di gunakan dalam penelitian. Dalam penelitian ini yang akan dicari adalah pengaruh kota tempat tinggal dengan kemampuan berbahasa Inggris, sehingga data alamat yang di dapat akan di transformasi sampai data yang tersisa adalah alamat kota saja.

3.2.4. Profiling Data Latitude dan Longitude Kota

Teknik yang digunakan pada pencarian ini adalah *data profiling*. *Profiling* yang dilakukan adalah mencari koordinat garis lintang dan bujur setiap daerah. Pencarian *longitude* dan *latitude* akan menggunakan *google maps*. Koordinat yang sudah ditemukan akan dimasukkan ke dalam *file excel*.

3.2.5. Pencarian Jarak Antar Kota dan Random sampling Data

Pencarian jarak antar kota dilakukan dengan menggunakan metode *Haversine*. Metode ini digunakan untuk menghitung jarak antar titik di permukaan bumi yang mengacu pada garis lintang atau *longitude* dan pada garis bujur atau *latitude*.

Haversine formula adalah persamaan yang memberikan jarak lingkaran besar antara dua titik berdasarkan garis bujur dan lintang. Dengan mengasumsikan bahwa bentuk permukaan bumi adalah bulat sempurna dengan jari-jari R 6.367,45 km, dan lokasi dari kedua titik di kordinat bulat sempurna itu adalah lon1, lat1, dan lon2, lat2. Dari hal ini rumus *haversine* dapat ditulis dengan persamaan sebagai berikut:

$$d = 2r \arcsin(\sqrt{\text{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1) \cos(\varphi_2) \text{hav}(\lambda_2 - \lambda_1)})$$

$$= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$

Rumus 3.1 Rumus *Haversine law*

Dimana:

φ = garis lintang atau *latitude*

λ = garis bujur atau *longitude*

d = jarak

Random sampling yang dijalankan pada penelitian ini digunakan untuk mengambil data secara acak pada data nilai mahasiswa UMN pada daerah Jakarta dan Tangerang. Metode *Random sampling* digunakan juga karena

hasil pengolahan data yang diluar batas kemampuan *Microsoft Excel* dan juga *Google Spreadsheet* untuk menampung semua hasil olahan data tersebut.

Tabel 3.2 Tabel perbandingan *tools random sampling*

<i>Tools/Software</i>	<i>Tableau</i>	<i>Power BI</i>
<i>Features</i>		
<i>Business Intelligence</i>	✓	✓
<i>Data Visualization</i>	✓	✓
<i>Dashboard Creation</i>	✓	✓
<i>Customize Reporting</i>	✓	✓
<i>Random sampling</i>	✓	✓
<i>Gartner Magic Quadrant</i>	✓	✓
<i>Subscription Free</i>	✓	✓
<i>Rating</i>	95%	98%

Pada penelitian ini *tools Random sampling* yang digunakan adalah *Tableau*. *Tableau* merupakan aplikasi bisnis intelejen yang mudah digunakan dalam hal membuat visualisasi data, analisis data, dan pelaporan. Cara menggunakannya mudah dengan sistem *drag and drop*. Pemilihan menggunakan *tools Tableau* juga dikarenakan memiliki akses untuk *tableau license*. *Tableau license* memudahkan untuk menggunakan *tools* dengan maksimal sehingga data yang dibaca pada *tableau* dapat lebih dari 1000 data.

3.2.6 Pencarian Pengaruh Bahasa Inggris dengan Jarak Geografis

Proses pengujian yang dilakukan akan menggunakan data yang telah dikumpulkan melalui proses sebelumnya. Pengujian akan menggunakan metode *linear regression* dengan bantuan *library* yang tersedia.

Model perhitungan yang digunakan di dalam penelitian skripsi ini adalah *Linear Regression* dengan *Ordinary Least Square*. Model ini dijadikan sebagai perhitungan dalam menentukan ada atau tidaknya sebuah keterkaitan atau pengaruh di antara letak kota dengan pemahaman bahasa Inggris dalam melihat *p-value*, dengan hasil nilai *R square* tertentu dari hasil penelitian. Pemilihan linear regression dilakukan berdasarkan jurnal penelitian sebelumnya yang melakukan perbandingan jarak geografis dengan relasi perbedaan lagu daerah.

3.2.7 Mencari Variansi Data Bahasa Inggris Antara Daerah

Setelah melakukan pencarian linear dari data ujian TOEIC mahasiswa UMN, selanjutnya akan mencari variansi data antar daerah. Tujuan dari hal ini adalah untuk mencari adakah variansi data yang dimiliki antara daerah tersebut. Pencarian akan menggunakan metode Analysis of Variance atau Anova. Metode Anova adalah sebuah analisis statistik yang menguji perbedaan rerata antara group. Hasil akhir dari tes Anova ini adalah untuk menemukan nilai p-value yang nantinya akan dibandingkan dengan nilai p-value pada table, jika p-value lebih dari p-value di table maka dapat disimpulkan bahwa H_0 ditolak dan H_1 diterima. Hal ini akan menyatakan kalau tidak terdapat keterkaitan antara daerah dengan pemahaman berbahasa Inggris dengan data ujian TOEIC mahasiswa UMN.