PREDIKSI LAJU PERTUMBUHAN JUMLAH PENDUDUK PROVINSI KALIMANTAN SELATAN MENGGUNAKAN METODE K-NEAREST NEIGHBOR REGRESSION

Hoiriyah[1], Fakhriani Ekawati[2]

[1]Fakultas Teknologi Informasi, Universitas Islam Kalimantan MAB Banjarmasin
Email: [1]ihaybjm18@gmail.com
[2]Fakultas Teknologi Informasi, Universitas Islam Kalimantan MAB Banjarmasin
Email: [2]fakhrianiekawati@gmail.com,

Abstrak

Penduduk adalah mereka yang berdomisili kurang dari 6 bulan tetapi bertujuan untuk menetap, jumlah penduduk pada suatu wilayah akan terus bertambah setiap tahunnya, di mana pertumbuhan tersebut dapat memberikan dampak baik maupun buruk sehingga di perlukannya wadah untuk menampung data pertumbuhan itu. Dengan adanya Data mining sebagai proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.ada banyak metode yang bisa di gunakan dalam menganalisanya salah satunya adalah Metode K-Nearest Neighbor Regression (KNN) merupakan algoritma yang digunakan untuk melakukan regresi atau peramalan untuk di gunakan sebagai prediksi dalam lajunya pertumbuhan penduduk di indonesia terutama yang ada di kalimantan selatan. Data pada penelitian ini merupakan data yang di gunakan selama 9 Tahun yang mana pada data yang ada pada kabupaten di sini di ubah terlebih dahulu kedalam binirisasi menggunakan Metode One Hot Encoding digunakan untuk merepresentasikan data bertipe kategori atau dalam pengkodingan dekenal dengan data bertipe string atau huruf menjadi data vektor biner yang bernilai integer, 0 dan 1, sehingga data bisa di proses menggunakan linear regression agar bisa di analisa lebih lanjut dengan menggunakan data label. Singga dari hasil analisa data yang di gunakan di atas dapat di simpulkan hasil dari Mean Squared Error " MSE " terbaik ada pada data label pada tahun 2012 dengan menggunakan data training 20% dan data testing 80% dengan memperoleh hasil Akurasi sebanyak 1.266 +/- 0.000 lebih baik dari hasil yang lain setelah di lakukannya beberapa kali analisa.

Keywords *: Penduduk, Datamining, K-Nearest Neighbor Regression (KNN)*

1. PENDAHULUAN

Dari data (Adminduk) Administrasi Kependudukan per Juni 2021, data penduduk di Indonesia adalah sebanyak 272.229.372 jiwa, dimana Permasalahan kependudukan telah menjadi masalah penting bagi pemerintah dan para pakar kependudukan di Indonesia[3]. Badan Pusat Statistik (BPS) tanggal 21 Januari 2021 telah merilis hasil survei kependudukan 2020. Diketahui data survei penduduk 2020 bahwa data penduduk Indonesia mulai September 2010 sejumlah 270,20 juta jiwa atau bertambah 32,56 juta perpenduduk yang memiliki pertambahan Kontribusi penduduk paling banyak yang mencapai lebih dari 5,25 juta jiwa berada di daerah Jawa Barat, setelahnya Jawa Tengah adalah 4,13 juta jiwa, dan Jawa Timur yaitu 3,18 juta [1]. Dari Hasil Sensus Penduduk 2020 mencatat kalimantan selatan jumlah penduduk mencapai 4,074 juta jiwa. Rinciannya, sebanyak 2,06 juta penduduk yang jenis kelamin laki-laki, sedangkan 2,01 juta jiwa perempuan. Kepadatan penduduk di Kalimantan Selatan mencapai 105 jiwa per kilometer persegi (km^2). Rata-rata laju pertumbuhan penduduk Kalimantan

Selatan sebesar 1,13% dalam satu dekade terakhir. Berdasarkan wilayahnya, Kota Banjarmasin memiliki jumlah penduduk terbanyak di Kalimantan Selatan, yakni 657,66 ribu jiwa. Sedangkan, Kabupaten Balangan paling sedikit jumlah penduduknya, yakni 130,36 ribu jiwa.

Pada dasarnya suatu wilayah akan selalu bertambah pada setiap tahunnya, yang mana pertumbuhan tersebut bisa berdampak baik dan buruk untuk suatu wilayah tersebut. Pertumbuhan penduduk jika ditangani dengan tepat dapat meningkatkan kualitas suatu wilayah, di antara lain dalam segi ekonomi, di mana pendapatan per kapita akan bertambah dengan meningkatnya jumlah penduduk dan juga dalam segi inovasi, yaitu berkembangnya inovasi-inovasi baru akibat dorongan pertumbuhan penduduk dari suatu wilayah sehingga di perlukannya wadah untuk menampung data pertumbuhan itu. Dengan adanya Data mining sebagai proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. " KDD " Knowledge Discovery in Database merupakan salah satu cara untuk memproses data sebelum menentukan metode atau algoritma yang sesuai dengan database yang akan di gunakan, algoritma yang tepat merupakan penentu untuk mendapatkan hasil yang terbaik secara keseluruhan[2] ada banyak metode yang bisa dipakai untuk menganalisanya salah satunya adalah Metode K-Nearest Neighbor Regression (KNN) merupakan algoritme yang sering dipergunakan untuk melakukan regresi atau peramalan untuk di gunakan sebagai prediksi dalam lajunya perkembangan data penduduk di wilayah indonesia terutama daerah di kalimantan selatan.

Untuk mengatasi permasalahan yang telah diuraikan di atas, maka peneliti akan melakukan penelitian dengan data berfokus pada penduduk daerah kalimantan selatan.

2. METODE PENELITIAN

2.1 Data Penelitian

Data yang digunakan pada penelitian ini berupa data penduduk daerah Kalimantan Selatan selama 9 Tahun mulai Tahun 2011 Hingga Tahun 2019 di peroleh dari BPS Provinsi Kalimantan Selatan.

Kabupaten	Laju Pertumbuhan Penduduk per Tahun (Persen)		
	2011	2012	2013
KALIMANTAN SELATAN	1,97	1,93	1,84
TANAH LAUT	1,81	1,75	1,69
KOTABARU	2,00	1,93	1,90
BANJAR	1,79	1,73	1,73
BARITO KUALA	1,57	1,48	1,54
TAPIN	1,59	1,51	1,49
HULU SUNGAI SELATAN	1,25	1,25	1,24
HULU SUNGAI TENGAH	1,40	1,29	1,26
HULU SUNGAI UTARA	1,47	1,45	1,30
TABALONG	1,82	1,80	1,76
TANAH BUMBU	4,75	4,48	3,78
BALANGAN	1,96	1,79	1,78
KOTA BANJARMASIN	1,54	1,48	1,40
KOTA BANJAR BARU	2,12	2,12	2,12

Sumber: Provinsi Kalimantan Selatan 2018-2019

2.2Data Mining

merupakan teknik pengolahan data yang penerapannya bisa digunakan di semua bidang, metode ini memiliki aturan agar sebelum di gunakan data harus disesuaikan terlebih dahulu sebelum digunakan, Ada tujuh klasifikasi data mining yang dibedakan berdasarkan cara kerjanya, seperti berikut ini..

1. Sequencing / Tracking Patterns

Langkah awal yang di gunakan ialah mencari pola atau mengurutkan kejadian sesuai dengan peristiwa. yang fungsinya menentukan jenis atau pola pada urutan kejadian (sequence) yang teratur. Teknik tracking patterns dapat menemukan pola interval tertentu, misalnya persentasi permintaan suatu produk saat akhir pekan atau data pengunjung situs pada musim tertentu..

2. Regression

Regression adalah metode Klasifikasi data mining yang berfungsi untuk mencari dan mengenali pola data yang bernilai numerik dan menentukan kelasnya, hasil cara ini diguanakan sebagai fungsi penentu untuk dijadikan dasar penginputan nilai, seperti, menggunaan salah satu harga suatu produk yang di dasarkan dari beberapa faktor tertentu dari produk yang telah ada, kompetitor, pelanggan dan permintaan.

3. Prediction / Forecasting

Tahapan terakhir data mining disebut sangat valuable dikarenakan tujuaanya untuk mengukur hasil prediksi nilai yang di peroleh pada suatu periode tertentu. dengan metode prediksi ini, datapat mendeteksi data ganguan "noise" yang ada di periode yang lalu di jadikan dasar dari prediksi sebagai acuaanya. contoh, berdasarkan data penjualan bulan sebelumnya, maka bisa

mempredikisi hasil penjualan yang akan terjadi di bulan yang selanjutnya.

2.3 K-NEAREST NEIGHBOR (KNN)

K-Nearest Neigbhor (KNN) merupakan salah satu metode yang populer untuk menghasilkan klasifikasi teks [6] yaitu dengan melakukan proses pembelajaran dari data latih untuk menentukan kelompok k objek. Sehingga dalam menentukan hasil klasifikasi KNN melihat jarak terdekat dari objek dengan masing-masing kelompok. Jarak tersebut diperoleh dari hasil kedekatan antara data masukan dengan data yang berada dalam kelompok berdasarakan nilai sejumlah fitur yang ada. Tetapi KNN juga memiliki kekurangan dalam aspek komputasi perhitungan salah satunya yang besar (Claudio, Fauzi, M. & Perdana, R. 2017) Tahapan untuk menjalani proses KNN sebagai berikut:

a. Menentukan jumlah pada tetangga k.
b. Menghitung jarak objek dengan masing-masing data kelompok. Perhitungan jarak menggunakan rumus euclidian distance (Sreemathy dan Balamurungan, 2012) yang ditunjukan pada persamaan 2.1

$$D(x,y) = \sqrt{\sum_{k=1}^{n}(x_i - y_i)^2} \qquad (2.1)$$

Dimana D adalah jarak, x dan y adalah data latih dan data uji.

2.4 *Mean Absolute Error (MAE) Atau Mean Absolute Deviation(MAD)*

Deviasi atau Simpangan secara mutlak diguankan untuk menentukan rata - rata data pusat sedangkan " MAD" Mean Absolute Deviation merupakan data peramalan yang di pergunakan untuk mengmenghitung tracking signal. penyebaran data bisa di lacak dengan singal yang digunakan untuk mencek apakah teknik prediksi peramalan bisa di pakai atau tidak. pada analisa inimenggunakan Mean Absolute Deviation (MAD) pada metode peramalan Moving Averages..

2.5 Mean Squared Error (MSE)atau Mean Squared DeviTION (MSD)

Mean Absolute Deviation (MAD) perhitungan yang digunakan untuk menghitung rata-rata kesalahan mutlak atau absolut. Mean Mean

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

absolute deviation (MAD) di pakai untuk mencari

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

akurasi dari forecast untiuk meratakan dari besarnya kesalahan yang terjadi pada setiap bagian forecasting yang bernilai absolut untuk masing - masing errornya.metode ini di pergunakan jika data analisa yang akan digunakan untuk mencek kesalahan peramalan dalam ukuran tertentu agar sama seperti data yang original. MAD mengukur ketepatan prediksi atau ramalan dengan mengurangi rata -rata kesalahan perkiraannya.

3. HASIL DAN PEMBAHASAN

Hasil Penelitian Prediksi Laju Pertumbuhan Jumlah Penduduk Provinsi Kalimantan Selatan Menggunakan Metode K-Nearest Neighbor Regression

1.Analisa untuk mendapatkan hasil MSE yang dilakukan dengan menggunakan data 2011 sebagai label dengan mengunakan data training sebanyak 90% dan testing sebanyak 10% yang memiliki nilai akurasi terbaik di banding dengan data training dan testing lainnya yang di gunakan dari label data 2011 atau bisa di lihat melalui uraian pada tabel 3.1 berikut ini.

Training	*Testing*	*Label*	*MSE*
90%	10%	2011	0.239 +/- 0.000
80%	20%		0.182 +/- 0.000
70%	30%		0.221 +/- 0.000
60%	40%		0.230 +/- 0.000

Tabel 3.1 tabel hasil Analisa MSE dari label 2011

2. Analisa untuk mendapatkan hasil MSE yang dilakukan dengan menggunakan data 2012 sebagai label dengan mengunakan data training sebanyak 20% dan testing sebanyak 80% yang memiliki nilai akurasi terbaik di banding dengan data training dan

testing lainnya yang di gunakan dari label data 2012 atau bisa di lihat melalui uraian pada tabel 3.2 berikut ini.

Training	Testing	Label	*MSE*
90%	10%	2012	0.077 +/- 0.000
80%	20%		0.070 +/- 0.000
70%	30%		0.870 +/- 0.000
60%	40%		0.945 +/- 0.000
50%	50%		0.987 +/- 0.000
40%	60%		1.055 +/- 0.000
30%	70%		1.215 +/- 0.000
20%	80%		1.266 +/- 0.000

Tabel 3.3 tabel hasil Analisa MSE dari label 2012

3. Analisa untuk mendapatkan hasil MSE yang dilakukan dengan menggunakan data 2013 sebagai label dengan mengunakan data training sebanyak 20% dan testing sebanyak 80% yang memiliki nilai akurasi terbaik di banding dengan data training dan testing lainnya yang di gunakan dari label data 2013 atau bisa di lihat melalui uraian pada tabel 3.3 berikut ini.

Training	*Testing*	*Label*	*MSE*
90%	10%	2013	0.046 +/- 0.000
80%	20%		0.048 +/- 0.000
70%	30%		0.673 +/- 0.000
60%	40%		0.716 +/- 0.000
50%	50%		0.742 +/- 0.000
40%	60%		0.791 +/- 0.000
30%	70%		0.908 +/- 0.000
20%	80%		0.948 +/- 0.000

Tabel 3.3 tabel hasil Analisa MSE dari label 2013

4. Analisa untuk mendapatkan hasil MSE yang dilakukan dengan menggunakan data 2014 sebagai label dengan mengunakan data training sebanyak 50% dan testing sebanyak 50% yang memiliki nilai akurasi terbaik di banding dengan data training dan testing lainnya yang di gunakan dari label data 2014 atau bisa di lihat melalui uraian pada tabel 3.4 berikut ini.

Training	*Testing*	*Label*	*MSE*
90%	10%	2014	0.028 +/- 0.000
80%	20%		0.030 +/- 0.000
70%	30%		0.499 +/- 0.000
60%	40%		0.531 +/- 0.000
50%	50%		0.547 +/- 0.000
40%	60%		0.531 +/- 0.000
30%	70%		0.499 +/- 0.000
20%	80%		0.030 +/- 0.000

Tabel 3.4 tabel hasil Analisa MSE dari label 2014

4. KESIMPULAN

Data yang di gunakan selama 9 Tahun mulai Tahun 2011 Hingga Tahun 2019 pada data yang di olahah data kabupaten di sini di ubah terlebih dahulu kedalam binirisasi menggunakan Metode One Hot Encoding digunakan untuk merepresentasikan data bertipe kategori atau dalam pengkoding dekenal dengan data bertipe string atau huruf menjadi data vektor biner yang bernilai integer, 0 dan 1, sehingga data bisa di proses menggunakan linear regression sehinga bisa di analisa lebih lanjut dengan menggunakan label dari data tertentu yang sudah ada.

Singga dari hasil analisa data yang di gunakan di atas dapat di simpulkan hasil dari Mean Squared Error " MSE " terbaik ada pada data label pada tahun 2012 dengan menggunakan data training 20% dan data testing 80% dengan memperoleh hasil Akurasi sebanyak 1.266 +/- 0.000 lebih baik dari hasil yang lain setelah di lakukannya beberapa kali analisa.

5. REFERENSI

[1] Hasil Survei Penduduk 2020 Peluang Indonesia Maksimalkan Bonus Demografi Diakses Pada Januari 2022, dari https://www.kemenkopmk.go.id/hasil-survei-penduduk-2020-peluang-

indonesia-maksimalkan-bonus-demografi

[2] Yuli Mardi, "Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," JurnalEdikInformatikaPenelitian Bid. Komput. Sains dan Pendidik. Inform., vol. 2, pp. 213–219, 2016.

[3] Badan Pusat Statistik (BPS), Diakses Pada Januari 2022, dari https://www.bps.go.id/subject/12/kependudukan.html

[4] Distribusi Penduduk Indonesia Per Juni 2021: Jabar Terbanyak, Kaltara Paling Sedikit Diakses Pada Januari 2022, dari https://dukcapil.kemendagri.go.id/berita/baca/809/distribusi-penduduk-indonesia-per-juni-2021-jabar-terbanyak-kaltara-paling-sedikit

[5] Han, J., Kamber, M. & Pei, J., 2012. Data Mining: Concepts and Techniques. 3rd ed. Waltham: Morgan Kaufmann.

[6] Sreemathy, J. dan Balamurungan, P. S., 2012. An efficient text classification using KNN and naïve bayesian. International Journal on Computer Science and Engineering (IJCSE), vol. 4, no. 3, pp. 392-396.