BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

Pada bab ini, akan dilakukan pembahasan mengenai sejumlah studi sebelumnya yang terkait dengan bidang Machine Learning dan Data Mining. Penelitian-penelitian ini menjadi acuan penting dalam penelitian saat ini yang bertujuan untuk menganalisis pemetaan tingkat kesejahteraan sosial di wilayah Jakarta.

2.2 *Machine Learning*

Bidang dalam kecerdasan buatan (*Artificial Intelligence, AI*) yang berfokus pada pengembangan algoritma dan teknik yang memungkinkan komputer untuk belajar dari dan membuat prediksi berdasarkan data. Secara umum, machine learning memungkinkan mesin untuk melakukan tugas-tugas seperti klasifikasi, regresi, clustering, forecasting dan lain-lain tanpa harus diprogram secara eksplisit untuk setiap tugas. Jenis-jenis Machine Learning seperti berikut :

1. Supervised Learning

Algoritma dilatih menggunakan data yang diberi label. Model belajar untuk memetakan input ke output yang benar berdasarkan contoh data pelatihan. Contoh metode termasuk regresi linear, decision tree, dan neural networks.

2. *Unsupervised Learning*

Algoritma dilatih menggunakan data yang tidak diberi label dan berusaha menemukan struktur atau pola dalam data. Contoh metode termasuk clustering dan association.

3. *Semi-supervised Learning*

Kombinasi dari supervised dan unsupervised learning, di mana sebagian data diberi label dan sebagian lainnya tidak. Metode ini berguna ketika pelabelan data yang memakan waktu.

4. *Reinforcement Learning*

Algoritma belajar dengan melakukan aksi di lingkungan tertentu dan menerima umpan balik dalam bentuk reward atau punishment. Metode ini sering digunakan dalam aplikasi seperti game dan robotik. (Iskandar dkk., 2022).

2.2.1 *K-Means Clustering*

Algoritma K-Means clustering adalah salah satu teknik yang paling populer dalam analisis data dan pembelajaran mesin yang digunakan untuk mengorganisir data ke dalam kelompok-kelompok atau klaster berdasarkan kemiripan karakteristiknya. Metode ini bekerja dengan cara berikut:

1. Inisialisasi

   Memilih k, jumlah klaster yang diinginkan. Secara acak menempatkan k titik pusat awal (centroid) di dalam ruang data.

2. Penetapan Klaster

   Setiap data di dalam kumpulan data akan dikaitkan dengan titik pusat terdekat. Hal ini dilakukan dengan menghitung jarak Euclidean antara data dan setiap titik pusat, kemudian data tersebut ditetapkan ke klaster dengan titik pusat terdekat.

3. Pembaruan Titik Pusat

   Setelah semua data ditetapkan ke klaster, titik pusat klaster diperbarui dengan menghitung rata-rata dari semua data dalam klaster tersebut. Titik pusat baru ini akan menjadi representasi sentral dari klaster.

4. Pengulangan

   Langkah penetapan klaster dan pembaruan titik pusat diulangi hingga tidak ada perubahan signifikan pada posisi titik pusat atau hingga jumlah iterasi yang telah ditentukan tercapai.

Kelebihan dari algoritma K-Means termasuk kesederhanaannya dan kemampuannya untuk menangani dataset yang besar. Namun, K-Means juga memiliki kelemahan, seperti kepekaan terhadap titik awal yang dipilih secara acak dan sulitnya menentukan jumlah klaster yang optimal.

Dengan menggunakan K-Means clustering, data yang memiliki kemiripan akan dikelompokkan bersama dalam satu klaster, sementara data yang berbeda akan ditempatkan dalam klaster yang berbeda. Hal ini membuat K-Means clustering menjadi metode yang efektif untuk mengelompokkan data yang serupa ke dalam klaster-klaster yang terpisah, sehingga memudahkan analisis lebih lanjut (Parama Yoga, 2023).

Tahap-tahapan dasar algoritma k-means diantaranya sebagai berikut : (Dwi Yuliyanti & Zuraidah, 2021).

1. Langkah pertama menetapkan pusat cluster yang dipilih secara random.
2. Menggunakan rumus persamaan jarak euclidean kemudian hitung setiap data responden ke pusat cluster random.

$$d(i, k) = \sqrt{\sum_i^m}(Cij - Ckj)\text{^}2 \quad (1)$$

3. Mengelompokkan data ke dalam cluster jarak yang paling kecil dengan persamaan :

$$\text{Min } \sum_k^i - aik = \sqrt{\sum_i^m}(Cij - Ckj)\text{^}2 \quad (2)$$

4. Lalu, menghitung pusat cluster baru dengan menggunakan persamaan

$$Ckj = \frac{\sum_k^i x ij}{p} \quad (3)$$

Dengan : Xij cluster ke K p= banyaknya anggota cluster ke – k

5. Mengulang langkah kedua sampai ke empat sampai sudak tidak ada lagi data yang berpindah ke cluster lain.

2.2.2 *ARIMA Forecasting*

Model ARIMA (AutoRegressive Integrated Moving Average), dikembangkan oleh Box dan Jenkins pada tahun 1970, adalah model yang serbaguna dan banyak digunakan untuk meramalkan data deret waktu. Model ini menggabungkan beberapa komponen untuk menangkap ketergantungan temporal dalam data. Komponen Autoregressive (AR) mewakili hubungan antara nilai saat ini dengan nilai-nilai sebelumnya (lag), sedangkan komponen Moving Average (MA) menggambarkan

pengaruh dari kesalahan ramalan pada periode sebelumnya. Selain itu, komponen "Integrated" (I) menunjukkan jumlah kali data harus di-diferensiasi untuk membuatnya stasioner.

Model ARIMA dinyatakan dalam tiga parameter: AR(p), yang menunjukkan orde dari model autoregressive; I(d), yang menunjukkan derajat diferensiasi yang diterapkan untuk membuat data stasioner; dan MA(q), yang menunjukkan orde dari model moving average. Kombinasi dari komponen-komponen ini memungkinkan ARIMA untuk menangkap berbagai pola dalam data deret waktu, membuatnya sangat efektif untuk berbagai aplikasi peramalan.

Secara matematis, model ARIMA dapat dinyatakan sebagai ARMA, yang menggabungkan komponen AR(p) dan MA(q) tanpa diferensiasi. Model ARMA mengasumsikan data sudah stasioner dan menggunakan kombinasi dari autoregressive dan moving average untuk memodelkan deret waktu. Formula dari model ARMA adalah:: (Alsuwaylimi, 2023)

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varepsilon_t - \varphi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (1)$$

Dimana ε_t adalah $\sim WN(0, \sigma^2)$, $\{\varphi = 1, \ldots, p\}$ and $\{\theta = 1, \ldots, q\}$ (2)

adalah koefisien AR(p) dan MA(q)

2.3 Kajian Penelitian Terdahulu

2.3.1 Paper 1 : Pemetaan Tingkat Kesejahteraan di Desa Tangaran Kabupaten Sambas Kalimantan Barat Menurut Indikator Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN), Rojia, Jurnal Global Futuristik, 2023, Sinta 5

Penelitian ini bertujuan untuk menggambarkan tingkat kesejahteraan keluarga di Desa Tangaran dengan menggunakan indikator yang telah ditetapkan oleh BKKBN. Selain itu, penelitian ini juga menganalisis faktor-faktor yang memengaruhi kesejahteraan keluarga, seperti pekerjaan, kondisi rumah tangga, dan kebiasaan pengeluaran. Penelitian ini memberikan pemahaman mendalam tentang distribusi tingkat kesejahteraan keluarga di Desa Tangaran, termasuk identifikasi masalah dan tantangan yang dihadapi oleh keluarga dalam mencapai kesejahteraan yang lebih baik. Di akhir penelitian, disajikan rekomendasi kebijakan atau tindakan yang dapat dilakukan untuk meningkatkan tingkat kesejahteraan keluarga di desa tersebut, dengan mempertimbangkan kondisi dan karakteristik masyarakat setempat.

Penelitian ini menggunakan pendekatan kuantitatif deskriptif, yang bertujuan untuk menggambarkan atau menginterpretasikan data dalam bentuk angka sesuai dengan realitas yang ada. Data lapangan dikumpulkan melalui kuesioner sebagai instrumen penelitian. Populasi penelitian adalah seluruh kepala keluarga di Desa Tangaran, yang berjumlah 1037 KK. Sampel penelitian ditentukan menggunakan rumus Slovin, yaitu sebanyak 43 KK, yang diambil secara proporsional dari populasi 1037 KK. Teknik pengambilan sampel dilakukan dengan cara Proportionate Stratified

Random Sampling, untuk memperhitungkan heterogenitas dalam populasi, dengan memperhatikan proporsi dari setiap dusun di Desa Tangaran. Data yang terkumpul dianalisis menggunakan metode statistik deskriptif untuk menggambarkan karakteristik dan distribusi tingkat kesejahteraan keluarga berdasarkan indikator yang telah ditetapkan oleh BKKBN.

Temuan utama dari penelitian ini menunjukkan bahwa mayoritas keluarga di Desa Tangaran berada pada tingkat pra-sejahtera. Sebagian besar pekerjaan di desa tersebut berkaitan dengan sektor pertanian. Terdapat perbedaan tingkat kesejahteraan antar dusun, di mana beberapa dusun memiliki tingkat kesejahteraan yang lebih tinggi dibandingkan dengan dusun lainnya. Faktor-faktor seperti ukuran rumah, pengelolaan pendapatan, dan partisipasi dalam kegiatan masyarakat diketahui memengaruhi tingkat kesejahteraan keluarga di Desa Tangaran.

2.3.2 Paper 2 : Optimalkan Peran Dinas Sosial Melalui Penggunaan Python Untuk Data Mining Pusdatin Kemensos Dalam Pelaksanaan Validasi Dan Verifikasi Dtks Di Provinsi Maluku, Syaefullah dkk, 2022, Sinta 5

Paper ini berfokus pada pengoptimalan peran Dinas Sosial melalui penggunaan Python untuk data mining di Pusdatin Kemensos dalam pelaksanaan validasi dan verifikasi Data Terpadu Kesejahteraan Sosial (DTKS) di Provinsi Maluku. Tujuan penelitian ini adalah untuk mengoptimalkan peran Dinas Sosial dalam memanfaatkan Python untuk proses data mining dan mengidentifikasi kendala atau hambatan yang dihadapi dalam upaya tersebut.

Penelitian ini menggunakan metode deskriptif kuantitatif, dengan mengumpulkan data dalam bentuk angka berdasarkan sampel besar dan data empiris yang disajikan. Data primer diperoleh melalui wawancara dengan Kepala Bidang Penanganan Fakir Miskin Dinas Sosial Provinsi Maluku, sementara data sekunder diperoleh dari DTKS. Penelitian ini dilakukan selama tiga minggu setelah pembukaan Praktek Lapangan III, dengan lokasi penelitian di Dinas Sosial Provinsi Maluku.

Temuan utama dari penelitian ini menunjukkan bahwa penggunaan Python di Pusdatin Kemensos memberikan manfaat signifikan dalam validasi dan verifikasi DTKS. Dinas Sosial memainkan peran penting dalam pengembangan kemampuan keluarga di garis kemiskinan, peningkatan tingkat kesejahteraan sosial, serta perancangan dan pelaksanaan kebijakan teknis. Selain itu, Dinas Sosial juga memberikan bimbingan teknis dalam perlindungan dan pemberdayaan sosial bagi masyarakat miskin. Namun, penelitian ini juga menemukan kendala yang dihadapi, termasuk tingkat kemahalan yang tinggi, kurangnya perhatian dari pemerintah daerah, serta masalah geografis dan transportasi yang mempengaruhi proses verifikasi dan validasi data.

2.3.3 Paper 3 : Penanggulangan Kemiskinan Melalui Pemberdayaan Sdm Dalam Mengelola Potensi Lokal Perdesaan, Mardi Murachman dkk, Jurnal Ilmu Pemerintahan Suara Khatulistiwa (JIPSK), 2021, Sinta 5

Penelitian ini bertujuan untuk menganalisis dampak kebijakan pemerintah terhadap kemiskinan di Kabupaten Musi Rawas, khususnya dalam konteks implementasi visi pembangunan daerah. Penelitian ini juga mengidentifikasi faktor-faktor yang

menyebabkan ketimpangan sosial dan ekonomi di Kabupaten Musi Rawas. Selain itu, penelitian ini menganalisis efektivitas dan relevansi kebijakan pembangunan yang telah dilaksanakan dalam mengatasi kemiskinan dan ketimpangan di daerah tersebut.

Metodologi yang digunakan dalam penelitian ini adalah gabungan dari pendekatan kualitatif dan kuantitatif. Data primer diperoleh melalui wawancara dengan berbagai pemangku kebijakan di Kabupaten Musi Rawas, termasuk Bupati, anggota Badan Pusat Statistik, dan kepala desa. Data sekunder diperoleh dari dokumen-dokumen resmi pemerintah terkait, seperti Rencana Pembangunan Jangka Menengah Daerah (RPJMD), serta data statistik dari Badan Pusat Statistik. Analisis data dilakukan menggunakan pendekatan deskriptif dan analisis konten.

Temuan utama dari penelitian ini menunjukkan bahwa kebijakan pembangunan yang telah dilaksanakan belum sepenuhnya efektif dalam mengatasi kemiskinan dan ketimpangan di Kabupaten Musi Rawas. Oleh karena itu, diperlukan peningkatan kualitas kebijakan pembangunan, termasuk pengarusutamaan gender, pemberdayaan masyarakat, dan peningkatan akses terhadap pendidikan dan kesehatan, untuk mengatasi kemiskinan dan ketimpangan secara lebih efektif di Kabupaten Musi Rawas.