

BAB II

LANDASAN LITERATUR

2.1. TINJAUAN TEORI

Beberapa teori dasar yang akan diambil untuk dijadikan bahan utama pengerjaan tesis ini adalah:

2.1.1. Data Mining

Salah satu langkah pengerjaan yang menggunakan beberapa data utama yang memerlukan respitori data yang besar untuk mengambil informasi yang penting dan valid dengan tujuan utama adalah untuk melakukan *extract data patterns* dan mendeteksi beberapa hubungan parameter di dalam data. Pengerjaan ini harus menghasilkan prediksi yang akurat supaya mempermudah peneliti membuat data visualisasi berbentuk *dashboard*.

Aplikasi penelitian metode *Data Mining* ini dikatakan sebagai educational data mining. Tujuan utama penggunaan *Educational Data Mining* ini adalah untuk mendevelop model tersebut sehingga bisa mengetahui performa dari setiap variabel yang tersedia. Metode-metode *Data Mining* ini lebih statistik dan terbaru (Križanić, 2020).

2.1.2. *Big Data*

Salah satu etimologi besar tentang karakteristik data yang merupakan bagian dari kualitas menganalisa dataset yang sangat besar dengan tiga karakteristik yang utama, yaitu volume, *velocity*, dan *variety*. Istilah ini banyak digunakan untuk mengatakan komponen-komponen dasar *Big Data* yang dikerjakan oleh para *Data Scientist*. (Kitchin & McArdle, 2016).

2.1.3. *Data Warehouse*

Komponen yang digunakan untuk melakukan pemrosesan data di dalam suatu perusahaan terkait dengan Big Data untuk membuat *Data Integration* menggunakan *framework* utama, yaitu *Extract-Transform-Load* (ETL). *Framework* ini memproses perhitungan suatu data. Contoh penerapan *Data Warehouse* terletak di bidang kesehatan. (Gavrilov et al., 2020).

2.1.4. *Data Visualization*

Metode pengerjaan untuk pembuatan *dashboard* presentasi yang dibuat secara semenarik mungkin untuk menarik perhatian customer berdasarkan hasil perhitungan matematis prediksi suatu permasalahan di sekitar. Contoh konkrit untuk melakukan data visualisasi ini terletak pada *tool* yang bernama MALDIviz, yaitu beberapa algoritma kompleks untuk membuat *dashboard* statistik pasien yang berada di rumah sakit karena akan membuat penelitian menjadi lebih menarik. (Jagadeesan & Ekström, 2017).

2.1.5. *Prediction*

Metode pengerjaan yang menggunakan penelitian survei yang terjadi di sekitar kita dengan tujuan untuk membuat perhitungan sebuah angka yang dapat memperkirakan suatu permasalahan yang dicatat dalam data yang valid. Contoh metode yang menggunakan *prediction* ini adalah *linear model* menggunakan *regression*. (Skeels & Taylor, 2015).

2.1.6. *Time Series*

Konsep *time series* yang digunakan terhubung dengan beberapa data prediksi yang akan dihitung dengan menggunakan model regresi untuk menentukan angka-angka data mentah dan akan dihitung berapa *error* yang didapat supaya keambiguan angka dapat diatasi. Metode ini diterapkan di pengerjaan *general error correction model* (GECM). (Enns et al., 2017).

2.1.7. *K-Medoids*

Algoritma model *K-Medoids* ini memiliki persamaan dengan model *K-Means*. Yang menjadi perbedaan antara model *K-Medoids* dengan model *K-Means* adalah model *K-Medoids* menggunakan objek perwakilan (*medoids*) sebagai pusat *clustering* di dalam sebuah penelitian *research*, sementara model *K-Means* hanya menggunakan nilai *mean* atau rata-rata sebagai pusat proses *clustering*. Algoritma model *K-Medoids* ini melengkapinya beberapa kelemahan yang terjadi di dalam model tersebut, yaitu terdapat suatu penyimpangan nilai value yang besar yang dapat mempengaruhi

pendistribusian data. Selain itu, proses *clustering* di dalam algoritma ini tidak bergantung dengan urutan dataset yang sudah dibuat (Pramesti et al., 2017).

2.1.8. DBScan

Salah satu model *clustering* yang dilakukan untuk mencluster suatu permasalahan tertentu dengan data yang tidak beraturan. Penelitian menggunakan model ini lebih efisien dibandingkan dengan *Clarans* karena model ini dapat mendeteksi *noise* atau *outlier* secara luas, tidak perlu adanya input seperti pada model *K-Medoids*, dan dapat mengenal beberapa data sampel yang tidak beraturan. Proses *clustering* model ini dapat dilakukan berdasarkan tingkat kedekatan/kerapatan jarak antar obyek dalam dataset (Sari & Primajaya, 2019).

2.1.9. X-Means

Sama seperti model *K-Means*, metode ini menyempurnakan limitasi atau kekurangan yang dimiliki oleh penelitian *clustering* model *K-Means* ini. Perbedaan yang sangat signifikan terdapat pada algoritma pengerjaan. Algoritma model ini akan menambahkan jumlah *cluster* yang ada secara rekursif sehingga jumlah yang dihasilkan sama dengan nilai user di atas. Penyelesaian sebuah penelitian menggunakan BIC daripada hanya satu iterasi pada model *K-Means*. (Wijayanto, 2019). dapat mengoptimalkan nilai BIC (*Bayesian Information Creation*). Model ini bisa memperlihatkan konsistensi terhadap nilai *value* yang dihitung (Heo et al., 2020).

2.1.10. Knowledge Discovery Of Database

Sebuah metode dimana semua pengetahuan dari suatu data yang terkait dengan sebuah organisasi ataupun lembaga yang memiliki jumlah data yang sangat besar dan berfokus kepada metodologi yang luas. Contoh penerapan ada di perhitungan Algoritma *FP-Growth* pada suatu pola yang ada di sekitar. (Informatika & Majapahit, 2019).

2.2. PENELITIAN TERDAHULU

Tabel 2.1. Penelitian Terdahulu

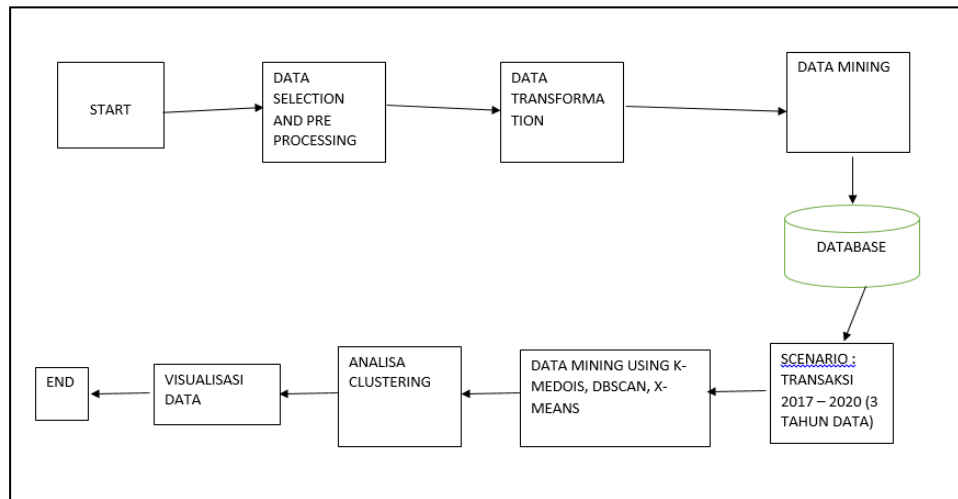
| Aspek permasalahan | Jurnal 1 | Jurnal 2 | Jurnal 3 |
|---|---|--|---|
| Nama peneliti | Friska Natalia Ferdinand, Yustinus Eko Soelistio, Ferry Vincentius Ferdinand, and I Made Murwantara | Hong-hai Zhang, Cheng-peng Jiang, dan Lei Yang | Todd Paciencia, Trevor Bihl, dan Kenneth Bauer |
| Tempat penelitian | Ahmad Dahlan University, Indonesia | College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, China | United States Air Force, 1570 Pentagon, Washington, DC |
| Bidang | <i>Information Technology</i> | Machine Engineering | Algorithms & Computational Technology |
| Nama Jurnal, Volume, Nomor Jurnal, Tahun | TELKOMNIKA, Vol.17, No.3, June 2019, pp.1376~1384 | Journal Advances in Mechanical Engineering, Vol. 8(9) 1–11, The Author(s) 2016 | Journal of Algorithms & Computational Technology Volume 13: 1–20 The Author(s) 2019 |
| Analisa jurnal | Jurnal ini memproses pengembangan <i>clustering</i> kondisi banjir di Tangerang berdasarkan | Melakukan evaluasi kemacetan lalu lintas untuk <i>forecasting Terminal Maneuvering Area</i> di Cina. Pendekatan penelitian ini menggunakan | Jurnal ini menggambarkan proses visualisasi <i>High-dimensional</i> data menggunakan 7 datasets sample <i>Mixed National Institute of Standards and</i> |

| | | | |
|-----------------------|--|--|--|
| | level ketinggian air yang berada di Angke, Pesanggrahan dan Cisadane menggunakan model <i>DBscan</i> , <i>K-Means</i> , dan <i>K-Medoids</i> . | metode penelitian kuantitatif. | <i>Technology</i> dengan model <i>Hyper-Radial Visualization</i> melalui pendekatan kualitatif. |
| Kesimpulan | Model <i>DBscan</i> , <i>K-Means</i> , dan <i>K-Medoids</i> ini bisa digunakan untuk melihat kewaspadaan masyarakat tentang kondisi banjir di Tangerang. | Pengerjaan metode ini sangat baik dilakukan dengan sampel yang sederhana dan telah membuat perkembangan yang sangat baik dan optimal untuk mengembangkan area <i>forecasting</i> di beberapa negara. Penelitian di paper ini bisa dilanjutkan dengan <i>traffic congestion awareness</i> berdasarkan <i>4D trajectory prediction</i> . | Model <i>HRV</i> ini bisa digunakan untuk menganalisa klasifikasi data yang kompleks dengan mempertimbangkan dimensi data yang sudah diberikan meskipun <i>subset class</i> terlalu besar karena mudah dan bisa divisualisasi. |
| Kritikan Studi | Hasil <i>output</i> yang dikerjakan mungkin tidak valid karena model yang digunakan sangat berbeda meskipun sudah berusaha digeneralisasi semua nilai yang dihasilkan. | Perhitungan periode <i>forecasting</i> sangat baik bila period $s = 1$, tetapi, level akan turun bila dinaikkan ke 30. | <i>Value</i> data yang di maksimalkan untuk korelasi <i>Hyper-radial Callculation (HRC)</i> membuat visualisasi menjadi lebih linear dan sulit untuk melihat karakteristik data dengan beberapa <i>objective</i> yang ada. |

2.3. KERANGKA KONSEPTUAL

Penelitian ini didasarkan kepada jurnal utama yang menjadi acuan utama penentuan faktor-faktor pengaruh volume penjualan di dalam perusahaan PT. Sinar Servindo Sakti, yaitu perhitungan prediksi nominal dan inventori barang oli sales selama 1 tahun dengan kategori oli hidrolik, oli

mesin, oli *compressor*, dan oli *gearbox*. Penggambaran konsep penelitian dapat diperlihatkan dalam gambar di bawah:



Gambar 2.1. Proses implementasi *clustering* di dalam data oli

Gambar 2.1 di atas memperhatikan beberapa langkah dasar yang dilakukan untuk mengerjakan penelitian thesis yang berhubungan dengan *forecasting* di berbagai kategori oli secara umum, seperti oli hidrolis, oli mesin, oli *compressor*, dan oli *gearbox*. Proses untuk perhitungan menggunakan metode KDD ini adalah data harus dipilih terlebih dahulu untuk melakukan *pre-process* data tersebut (4.1). Lalu, data tersebut ditransformasi untuk dijadikan data baru (4.2). Data tersebut akan dilakukan *data mining* untuk *database* (4.3). Tahapan ini hanya menggunakan satu skenario besar, yaitu menghasilkan volume penjualan sales barang di perusahaan. Model-model yang sudah ada akan melakukan proses *data*

mining untuk model (4.4) kemudian di lakukan proses *clustering* agar semua kelihatan jelas (4.5). Proses tersebut akan divisualisasikan untuk menghasilkan sebuah *dashboard* yang baik (4.6).

2.4. VARIABEL PENELITIAN

Berdasarkan konsep dan beberapa teori yang sudah dibuat, dengan demikian maka variabel penelitian yang akan menjadi perhatian untuk penelitian ini terletak di *forecasting* nominal dan inventori barang perusahaan PT. Sinar Servindo Sakti dengan kategori oli hidrolis, oli mesin, oli *compressor*, dan oli *gearbox*.