BAB II

LANDASAN TEORI

2.1 Pengertian Big Data Analytics dalam Sistem Informasi

Perkembangan teknologi yang semakin cepat hingga adanya persaingan teknologi antar negara. Perkembangan teknologi yang mewajibkan para generasi muda untuk saling mengembangkan kemampuannya sendiri agar tidak merasa terbelakang dengan perkembangan teknologi yang terjadi, supaya kemampuan manusia tidak tergantikan oleh mesin teknologi yang semakin berkembang dapat menggantikan posisi manusia dalam proses perekonomian negara.

Perkembangan teknologi membuat banyaknya fitur yang dipergunakan untuk saling bertukar pikiran, situs social untuk mencari dan mendapatkan teman dari berbagai daerah berbeda, serta mencari relasi untuk pekerjaan dan lain sebagainya. Adanya perkembangan teknologi ini membuat semakin banyaknya data yang harus ditampung oleh situs tersebut. Sebelum munculnya teknologi yang sekarang semakin canggih, pengolahan data selalu menjadi tugas seorang programmer dan pengolahan tersebut sangatlah memakan waktu lama.

Kumpulan data yang banyak dengan berukuran sangat besar disebut dengan *big data.* Big data adalah suatu perkembangan teknologi yang mencakup dalam dunia bisnis dan teknologi. Big data memiliki 5V yang adalah memiliki ukuran yang besar (*volume*), kecepatan (*velocity*), ragam (*variety*), memiliki nilai tertentu (*value*), dan berasal dari sumber yang akurat (*veracity*).

Adanya big data bertujuan agar dalam dunia bisnis, organisasi, ataupun individu dapat mengolah data agar mendapatkan informasi yang lebih jelas dan mendalam (*insights*) yang membuat pengambilan keputusan secara adil (*decision making*) dengan sebuah tindakan yang benar untuk kepentingan bersama (*action*) [3].

2.2 Kondisi Perekonomian Indonesia

Perekonomian Indonesia mengalami penurunan semenjak 4 tahun terakhir dari tahun 2019. Pengaruh yang negara Indonesia alami tidak lepas dari kondisi perlambatan pertumbuhan ekonomi di dunia yang hanya tumbuh 2,9 persen. Adapun faktor yang berdampak kepada perekonomian dunia dengan beberapa kejadian seperti negara Hong Kong, Iran,dan Irak yang mempengaruhi stabilitas ekonomi nasional [2].

Dari dalam negeri, masalah ekonomi yang terjadi mempengaruhi penurunan dari tahun sebelumnya, salah satunya terjadi pada sektor investasi. Meskipun terjadi penurunan di beberapa sektor, perekonomian Indonesia dapat stabil dengan adanya permintaan domestik dan dengan adanya inflasi yang tetap terjaga. Pada tahun 2019, dengan adanya pemilu turut dalam membantu perekonomian Indonesia. Penurunan yang terjadi pada perekonomian Indonesia adalah bagian primer dan sekunder, tetapi lain hal pada bagian tersier yang tumbuh menguat, diantaranya pada sektor bagian informasi dan komunikasi yang mengikuti cepatnya perkembangan teknologi di dunia, sektor jasa asuransi dan keuangan, serta sektor *real estate* yang dengan perkembangan teknologi ini dapat mempermudah dalam hal promosi dan penjualan via online dan memberikan banyak penawaran terbaik dalam penjualannya [2].

2.3 Ekspor

Hubungan yang dibentuk menjadi sebuah kerja sama antara perdangan intermasional dan pertumbuhan ekonomi menjadikan terhubungnya komunikasi antar negara. Proses perekonomian melalui ekspor adalah proses pengiriman dan penerimaan produk yang diproduksi oleh negara tersebut untuk dikirimkan antar negara supaya menambah penghasilan keuangan negara dan terbentuknya kerjasama antar negara di dunia.

Ekspor adalah proses pengiriman atau penjualan yang dilakukan oleh negara dari dalam negeri ke luar negeri [4]. Ekspor termasuk produk migas dan tanpa migas. Keuntungan yang didapatkan dari proses ekspor antar negara adalah bentuk

dari kerjasama antar negara di dunia dan dapat memperluas pasar, menambah devisa negara, serta memperluas lapangan kerja untuk daerah yang memproduksi keperluan untuk dikirimkan ke negara lain.

2.4 Metode Analisis Data

Metode penelitian ini menggunakan metode analisis data kuantitatif. Metode ini terfokus pada data perekonomian sebelum adanya virus Covid-19 yang terjadi di seluruh negara di dunia. Suatu kebijakan pemerintah yang memberlakukan *lockdown* membuat hambatan kegiatan ekonomi di setiap negara, begitu juga dengan kegiatan ekspor di setiap negara yang sudah menjadi kegiatan sebagai penambah pemasukan ke dalam negara. Dengan menggunakan data ekspor yang tercatat dari tahun 2014 hingga 2020 yang dapat dipergunakan untuk memprediksi tingkatan perekonomian selanjutnya jika dalam skala perekonomian masih turun karena adanya Covid-19.

2.5 Algoritma *Time Series*

Algoritma *time series* adalah suatu analisa data yang terfokus kepada waktu. Teknik yang digunakan untuk perhitungan data deret waktu, agregasi, dan peramalan untuk suatu pemodelan dengan deret waktu yang telah terdata dalam sebuah dataset [5].

2.5.1 Algoritma *ARIMA*

ARIMA adalah singkatan dari *Autoregressive Integrated Moving Average*. Biasanya model *ARIMA* digunakan untuk meramalkan deret waktu.

Model *ARIMA* ditentukan oleh tiga parameter urutannya, p, d, q. p menentukan jumlah istilah *Autoregressive* dalam model. d menentukan jumlah perbedaan yang diterapkan pada nilai deret waktu. q menentukan jumlah istilah *Moving Average* dalam model.

Model *ARIMA* adalah model *ARMA* terintegrasi. Model *ARMA* mendeskripsikan deret waktu dengan jumlah tertimbang dari nilai deret waktu tertinggal (istilah *Autoregresif*) dan jumlah tertimbang residual tertinggal. Residu

ini berasal dari proses kebisingan terdistribusi normal. "Terintegrasi" menunjukkan bahwa nilai model *ARMA* terintegrasi, yang sama dengan nilai deret waktu asli yang dijelaskan oleh model *ARMA* dibedakan.

Operator *ARIMA* menyesuaikan model *ARIMA* dengan diberikan p, d, q ke deret waktu dengan menemukan koefisien p + q (dan jika konstanta estimasi benar, konstanta) yang memaksimalkan *loglikelihood* bersyarat dari model yang menggambarkan deret waktu. Untuk optimasi digunakan algoritma *LBFGS* (*Limited-memory Broyden-Fletcher-Foldfarb-Shanno*).

Jika memilih nilai untuk p, d, q, penting bahwa *loglikelihood* bersyarat hanya merupakan estimasi yang baik untuk *loglikelihood* yang tepat jika jumlah parameter (jumlah dari p, d, q) tidak sesuai dengan urutan panjang deret waktu. Oleh karena itu, jumlah parameter harus jauh lebih kecil daripada panjang deret waktu.

Seberapa baik model *ARIMA* yang terlatih menggambarkan deret waktu tertentu sering dihitung dengan *Akaikes Information Criterion* (*AIC*), *Bayesian Information Criterion* (*BIC*) atau *Akaikes Information Criterion* (*AICC*) yang dikoreksi. Operator *ArimaTrainer* menghitung ukuran kinerja ini dan mengeluarkan *Performance Vector* yang berisi nilai yang dihitung. Model *ARIMA* yang menggambarkan deret waktu dengan baik memiliki kriteria informasi yang kecil. Operator ini hanya bekerja pada deret waktu numerik [5].

2.6 Penelitian Terdahulu

Tabel 2.1 Penelitian Terdahulu

No.	Nama Peneliti	Judul Penelitian	Nama Jurnal	Hasil Penelitian
1.	Kiki Fatmawati, Agus Perdana Windarto	Data Mining: Penerapan *RapidMiner* dengan *K-Means Cluster*	*CESS (Journal of Computer Engineering System and*	Pengujian pada penelitian ini, iterasi *clustering* sebanyak 6 kali, lalu data diolah menggunakan *tools*

		pada Daerah Terjangkit Demam Berdarah Dengue (DBD) berdasarkan Provinsi	*Science*) Vol. 3 No. 2 Juli 2018	*RapidMiner* untuk menentukan nilai *centroid* dari 3 kluster, yaitu C1 untuk tingkat tinggi, C2 untuk tingkat sedang, dan C3 untuk tingkat rendah. Dari 34 provinsi, terdapat 4 kluster tingkat tinggi untuk penderita kasus DBD yaitu Provinsi Jawa di Indonesia dan Sumatera Utara. Lalu terdapat 13 kluster dengan tingkat sedang, dan 17 kluster dengan tingkat rendah.
No.	**Nama Peneliti**	**Judul Penelitian**	**Nama Jurnal**	**Hasil Penelitian**
2.	Billy Riadi, Johan Setiawan	Visualisasi Data dan Prediksi Perkembangan Impor Barang di Indonesia Tahun 2010 – 2014	ULTIMA InfoSys, Vol. IX, No. 2 | Desember 2018	Dalam penelitian ini menunjukan bahwa Indonesia setiap tahunnya membutuhkan barang impor untuk membantu perekonomian di indonesia, contohnya barang impor yaitu dalam hal pertambangan yang

				berarti Indonesia masih kurang dalam pengetahuan cara pengolahan hasil tambang. Maka dapat diketahui jika Indonesia mampu dalam hal pengolahan hasil tambang dapat mengurangi hasil impor yang mempengaruhi dalam penghematan devisa negara. Pada penelitian ini menggunakan metode *linear regression* untuk memprediksi dan mendapatkan tingkat akurasi yang tinggi untuk data yang digunakan dalam prediksi impor barang Indonesia, dengan rata – rata nilai akurasi mencapai 80-90% dan ada beberapa kategori dengan tingkat akurasi menurun ke angka 60-70%, dikarenakan faktor *dependent* yang digunakan hanya satu,

				sedangkan banyak hal yang dapat mempengaruhi sektor impor, contohnya dalam hal pengeluaran konsumsi, tingkat kurs mata uang, dan pendapatan negara.
No.	**Nama Peneliti**	**Judul Penelitian**	**Nama Jurnal**	**Hasil Penelitian**
3.	Stephen Mangara Wainana, Joseph Njuguna Karomo, Rachael Kyalo, Noah Mutai	*Using Data Mining Techniques and R Software to Analyze Crime Data in Kenya (*metode *CRISP-DM)*	*International Journal of Data Science and Analysis* 2020; 6(1): 20-31	Setelah menerapkan algoritma *K-Means* pada kumpulan data kejahatan 2015, kategori kejahatan seperti 'perampokan' dan 'pencurian' memiliki kelompok berbeda yang memiliki hubungan linier yang sangat kuat. Juga, jumlah penduduk dan kejahatan untuk tahun 2015 membentuk kelompok-kelompok yang sangat terkait. Ada juga kategori kejahatan tidak terkait erat dan mereka membentuk grup k yang tidak

				memiliki hubungan linier. Misalnya, 'korupsi' dan 'pencurian saham 'memiliki korelasi negatif yang rendah. Algoritma *APRIORI* menunjukkan bahwa banyak kejahatan yang terkait.
No.	**Nama Peneliti**	**Judul Penelitian**	**Nama Jurnal**	**Hasil Penelitian**
4.	M. Alfan Alfian Riyadi, Kartika Fithriasari, dan Dwiatmono	Data Mining Peramalan Konsumsi Listrik dengan Pendekatan *Cluster Time Series* sebagai *Preprocessing*	JURNAL SAINS DAN SENI ITS Vol. 5, No.1, (2016) 2337-3520 (2301-928X Print)	Pada penelitian ini menggunakan metode peramalan yang digunakan yaitu *ARIMA* dan *ANN* menggunakan studi kasus data *benchmark* konsumsi listrik di Portugal. Hasil yang didapatkan adalah sebanyak 7 kluster dengan anggota kluster terbanyak pada kluster ke-4 yaitu sebanyak 120 *client*. Lalu untuk model peramalan dengan metode *ANN* lebih baik dibandingkan *ARIMA*. Menghasilkan sebanyak 259 dari

				sebanyak 348 *client* yang menyatakan bahwa permodelan dengan menggunakan metode *ANN* lebih baik dibandingkan *ARIMA*.
No.	**Nama Peneliti**	**Judul Penelitian**	**Nama Jurnal**	**Hasil Penelitian**
5.	Adhitio Satyo Bayangkari Karno	Analisis Data *Time Series* Menggunakan *LSTM* (*Long Short Term Memory*) dan *ARIMA* (*Autocorrelation Integrated Moving Average*) dalam Bahasa *Python*	ULTIMA InfoSys, Vol. XI, No. 1 | Juni 2020	Dalam penelitian ini memprediksi saham Telkom yang mencapai tingkat akurasi sangat tinggi yaitu 99% (*RMSE*=1%), dengan pengolahan data dengan metode transformasi *loging* dan *EWMA* (*ts_log_ewma*) atau *loging* kemudian *Moving Average*. Hasil untuk memprediksi data saham ini memperlihatkan dengan metode *LSTM* lebih akurat dibandingkan dengan menggunakan model statistik *ARIMA*. Dalam memilih proses

				transformasi yang sesuai untuk model *ARIMA* dan *LSTM* diperlukan perkembangan agar memperolah hasil yang lebih baik lagi.

Berikut adalah penjelasan pada tabel 2.1 penelitian terdahulu, yaitu :

1. Pada penelitian terdahulu yang berjudul Data Mining: Penerapan *RapidMiner* dengan *K-Means Cluster* pada Daerah Terjangkit Demam Berdarah Dengue (DBD) berdasarkan Provinsi menjadi salah satu referensi dalam penelitian ini dikarenakan dalam pengolahan data menggunakan *tools RapidMiner*. Sebab, penelitian ini berhubungan dengan bidang *Big Data Analytics* dalam pengolahan *data mining* dengan menggunakan salah satu algoritma pada *tools RapidMiner*. Pada penelitian ini, operator yang digunakan berasal dari *tools RapidMiner* untuk menghasilkan nilai prediksi dengan algoritma *time series* yaitu *ARIMA*.
2. Pada penelitian terdahulu yang berjudul Visualisasi Data dan Prediksi Perkembangan Impor Barang di Indonesia Tahun 2010 – 2014 menjadi salah satu referensi yang dapat digunakan pada penelitian ini, dikarenakan jurnal tersebut menghasilkan visualisasi data dan prediksi dari perkembangan barang import di Indonesia. Sedangkan pada penelitian ini menghasilkan visualisasi data dan prediksi dari perkembangan barang ekspor di Indonesia menggunakan data pada tahun sebelumnya dengan menggunakan salah satu algoritma untuk melihat tingkat akurasi untuk bentuk data yang digunakan dalam prediksi.
3. Pada penelitian yang berjudul *Using Data Mining Techniques and R Software to Analyze Crime Data in Kenya* merupakan salah satu referensi jurnal internasional yang membahas tentang analisis suatu data di sebuah

negara dengan menggunakan konsep *data mining* yaitu proses *CRISP-DM*. Seperti hal nya dalam penelitian ini, bahwa proses pengolahan data menggunakan teknik *data mining* untuk analisa sebuah data di suatu negara dengan menggunakan salah satu algoritma.

4. Pada penelitian yang berjudul Data Mining Peramalan Konsumsi Listrik dengan Pendekatan *Cluster Time Series* sebagai *Preprocessing* menjadi salah satu referensi dalam penelitian ini, karena penelitian tersebut menggunakan metode yang sama dalam penelitian ini sebab melakukan peramalan *data mining* dengan menggunakan algoritma *time series* dalam pengolahan datanya pada suatu negara. Algoritma yang digunakan dalam pengolahan data pun sama yaitu dengan algoritma *time series ARIMA*.
5. Pada penelitian dengan judul Analisis Data *Time Series* Menggunakan *LSTM* (*Long Short Term Memory*) dan *ARIMA* (*Autocorrelation Integrated Moving Average*) dalam Bahasa *Python* menjadi referensi penelitian ini dikarenakan proses dalam menganalisis data yang dilakukan menggunakan cara yang sama dalam pemprosesan data *time series* dengan menggunakan algoritma *ARIMA*. Hasil yang didapatkan pun sama seperti penelitian ini yaitu untuk memprediksi dengan menggunakan data sebelumnya.