# BAB II

# LANDASAN TEORI

## **Tinjauan Studi**

Pada sebuah upaya dalam melakukan suatu analisis, maka dibutuhkan suatu panduan ataupun rujukan serta dukungan untuk setiap hasil analisis yang sudah ada sebelumnya. Yang tentunya panduan atau rujukan tersebut akan berkaitan dengan suatu analisis yang sedang dilakukan. Hasil dari penelitian-penelitian terdahulu tersebut terdiri dari topik dan pembahasan hasil penelitian yang telah dilakukan sebelumnya.

### **Penelitian Terkait**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Area Penelitian | Karakteristik Data | Metode | Hasil Penelitian |
| 1. | Gaji karyawan berdasar pengalaman lama bekerja.[7] | Data gaji pegawai dengan pengalaman lama bekerja, variabel gaji, variabel masa kerja. | *Machine Learning.* | Teknik model analisis data menggunakan model analisis regresi linear untuk pengetahui pengaruh variable bebas terhadap variable terikat. Dari hasil prediksi gaji karyawan berdasarkan pengalaman lama bekerja, terdapat pengaruh positif dan signifikan antara gaji dan masa kerja terhadap kinerja karyawan. |
| 2. | Harga saham perusahaan pelayaran PT. BULL.[9] | Data primer: informasi perusahaan. Data sekunder: data historis harga saham 2019-2020 dari Yahoo Finance dan Laporan Tahunan Bursa Efek Indonesia. *Purposive* *sampling*: perusahaan pelayaran PT. BULL. *Quota sampling: data time series* periode harian, variabel harga pembukaan, variabel harga tertinggi, variabel harga terendah, variabel harga penutupan, variabel volume saham selama 1 tahun 2 bulan (Juni 2019 – Juli 2020). | *Machine Learning.* | Penelitian ini menggunakan model *Cross Industry Standard Process for Data Mining* (CRISP-DM). Dari prediksi harga saham, hasil penelitian menunjukkan terdapat selisih antara harga penutupan saham luaran data testing dengan harga penutupan saham aktual yang ada di bursa saham. |
| 3. | Harga sembilan bahan pokok di DKI Jakarta.[10] | Data sembako DKI Jakarta (1 Januari 2016 – 31 Desember 2019). Variabel tanggal, variabel komoditas, variabel pasar, variabel harga. | *Machine Learning.* | Model yang digunakan dalam penelitian yaitu regresi linier berganda. Dari hasil prediksi yang telah dilakukan, persentase sumbangan pengaruh variabel bebas terhadap variabel terikat sebesar 84,2%, sedangkan sisanya sebesar 15,8% dimana dipengaruhi oleh variabel yang tak dimasukkan dalam penelitian ini. |
| 4. | Harga rumah dari 2 website jual beli rumah.[11] | Studi documenter *web scrapping* 2 website (1 Oktober 2020 – 31 Oktober 2020). Variabel luas lahan, variabel luas bangunan, variabel banyaknya kamar tidur, variabel banyaknya kamar mandi, variabel ketersediaan tempat parkir mobil. | *Machine Learning.* | Model dalam membuat prediksi adalah metode regresi linear. Dari hasil penerapan prediksi harga rumah, pengolahan awal data yang dilakukan pada data set 7442 data menjadi 794 data sangat mempengaruhi dalam tingkat akurasi prediksi harga rumah. Adapun hasil akurasi terbaik menghasilkan tingkat akurasi prediksi sebesar 0.8859830993050699 atau 88%. Namun, memiliki nilai galat cukup tinggi sebesar 259171.91 atau Rp. 259.171,91. |
| 5. | Pengadaan inventaris barang.[12] | Data Inventarsi Barang Dinas Pariwisata Pemuda dan Olahraga Kota Tasikmalaya 2012 – 2016. Variabel kode barang, variabel tahun, variabel harga barang. | *Machine Learning.* | Model yang digunakan dalam penelitian adalah metode regresi linier. Dari prediksi yang dilakukan, prediksi harga barang minimal yaitu sebesar 3011855.102, dan prediksi harga barang maksimal yaitu sebesar 23752745.511. Hasil akurasi RMSE adalah 0.934. |
| 6. | Prediksi kasus Covid-19 di Indonesia.[13] | Data yang dipakai dari databooks yang terdiri dari atribut total kasus, kasus sembuh, kasus meninggal dan kasus aktif, pada penelitian ini hanya menggunakan data kasus aktif. | *Machine Learning.* | Penelitian ini menggunakan model *backpropagation* dan regresi linear. Dari hasil simulasi Matlab dengan tiga fungsi pelatihan yaitu traincgb didapatkan nilai rata-rata error yaitu sebesar 0,017107, dengan fungsi pelatihan traingd didapatkan nilai rata-rata error sebesar -0,55116 serta fungsi pelatihan traingdx didapatkan nilai rata-rata error sebesar -3,82202 sehingga dapat disimpulkan fungsi pelatihan yang paling konvergen yaitu fungsi pelatihan terhadap traingdx denagn nilai rata-rata error paling kecil. |
| 7. | Prediksi Harga Emas.[14] | Dalam penelitian ini, data yang digunakan berasal dari internet. Kriteria atau variabel yang digunakan dalam penelitian ini, yaitu harga buka, harga beli (Input) dan harga jual (Output). | *Machine Learning* | Digunakan tiga model dalam penelitian ini, yaitu regresi linear, *backpropagation*, dan fuzzy mamdani. Hasil dari ketiga model menunjukkan bahwa korelasi dari regresi linear sangat bagus, yaitu 0,929. Dan nilai korelasi tertinggi dari ketiga metode berasal dari metode backpropagation. Hal ini terbukti bahwa dalam memprediksi harga emas menggunakan backpropagation lebih sedikit errornya ±0,05. |
| 8. | Prediksi Besaran Pendapatan Daerah.[15] | Data yang digunakan adalah data besaran pendapatan Kabupaten Deli Serdang tahun 2017 dan 2018, yang akan dibagi menjadi 2 bulan, Triwulan (3 bualn), Caturwulan (4 bulan), 1 semester (6 bulan) dan 1 tahun (12 bulan). | *Machine Learning* | Dalam penelitian ini model yang digunakan adalah regresi linear sederhana. Berdasarkan proses hasil perhitungan yang dilakukan, diketahui nilai prediksi besaran pendapatan daerah dinas pendapatan daerah Kab. Deli Serdang. Dengan menggunakan algoritma regresi linear sederhana, dinilai dapat memprediksi besaran pendapatan daerah dinas pendapatan daerah Kab. Deli Serdang untuk tahun berikutnya sehingga program-program yang telah direncanakan sebelumnya oleh dinas pendapatan dapat berjalan denagn lancar, dan juga dapat membuat program-program yang baru agar dapat meningkatkan pendapatan daerah untuk memajukan daerah tersebut. |
| 9. | Prediksi Tingkat Produksi Kopi.[16] | Data produksi kopi diperoleh dari BPS Kabupaten Manggarai dengan mengambil data produksi kopi lima tahun terakhir yaitu dari tahun 2011 – 2015. | *Machine Learning* | Dalam penelitian ini mode; yang digunakan adalah model regresi linear sederhana. Berdasarkan hasil penelitian dan pembahasan tentang prediksi tingkat produksi kopi menggunakan regresi linear sederhana maka dapat disimpukan bahwa data yang digunakan untuk memprediksi kopi merupakan data *time series*, dan setelah melakukan perhitungan prediksi pada tahun 2011 – 2015 nilai tertinggi pada tahun 2015 sebesar 1.537,38 ton dan nilai terendah pada tahun 2011 sebesar 1.109,944 ton. Setelah dilakukan pengujian menggunakan MSE dan MAPE diperoleh nilai MSE 43,112% dan MAPE 20,001% sehingga pengyjian menggunakan MAPE jauh lebih baik dalam menghitung akurasi prediksi produksi kopi. |
| 10. | Peramalan Penerimaan Mahasiswa Baru Universitas Samudra.[17] | Data didapat dari laman web Universitas Samudra. Data diambil dari data mahasiswa yang diterima 5 tahun sebelumnya. | *Machine Learning* | Model yang digunakan dalam penelitian adalah model regresi linear sederhana. Dapat disimpulkan bahwa penggunaan metode forecasting (peramalan) untuk melakukan peramalan jumlah PMB yang akan masuk ke Universitas Sambudra pada 1 tahun yang akan datang dapat dikatakan bermanfaat dan jumlah error yang didapat juga tidak terlalu besar sehingga peramalan ini dapat dipertimbangan untuk universitas agar pihak universitas dapat meningkatkan berbagai hal untuk mempertimbangkan apa-apa saja yang yang harus ditingkatkan agar penerimaan mahasiswa baru pada tahun selanjutnya dapat meningkat. |
| 11. | Prediksi gaji menggunakan teknik regresi. [18] | Data *position, level, salary.* Variabel *position*, variabel *level*, dan variabel *salary*. | *Machine Learning* | Model yang digunakan adalah model regresi linear dan regresi polinomial. emilih gaji dari grafik x-y membutuhkan banyak percobaan karena mungkin ada lebih dari satu grafik yang sesuai. Prediksi ini benar hingga waktu tertentu. Akurasi dapat diperoleh dengan menerapkan regresi k-terdekat. |
| 12. | Analisis empiris teknik regresi berdasarkan harga rumah  dan prediksi gaji. [19] | Variabel pada dataset *employee* *salary : salary, total years of experience, certification, lead.* Variabel pada dataset *house prices : price, sqft\_living, bedrooms, bathrooms, sqft\_living15, sqft\_lot, sqft\_above, sqft\_basement.* | *Machine Learning* | Model yang digunakan adalah *Simple Linear Regression* (SLR) dan *Multiple Linear Regression* (MLR). *Multiple Linear Regression* baik daripada *Simple Linear Regression*. Karena pada data House price, MLR memiliki R-Square 0,67 dan SLR 0,49. Serta pada data prediksi gaji, MLR memiliki R-Square 0,92 dan SLR 0,75. |
| 13. | Analisis korelasi antara gaji dan efisiensi inovasi enterprise berdasarkan psikologi entrepreneur.[20] | Data perusahaan manufaktur A-share China dari tahun 2012 -2016. *Explanatory variable : Monetary salary of senior manager, Senior management salary, Ordinary employee salary, Salary gap within the senior management team, Salary gap between senior managers and ordinary employees. Explained variable : Enterprise innovation efficiency. Control variable : Enterprise size, Return on assets, Asset-liability ratio, Years of establishment, Dummy variable*. | *Machine Learning* | Metode yang digunakan dalam penelitian ini yaitu *multiple regression. Monetary salary* dan *equity salary of senior managers, ordinary employee salary,* dan *the internal salary gap of the senior management team* semuanya berkorelasi positif secara signifikan pada tingkat 1% dengan efisiensi inovasi perusahaan. Korelasi antara *salary gap* antara *senior managers* dan *ordinary employees* dengan *enterprise innovation efficiency* tidak jelas. Koefisien korelasi regresi antar variabel berada dalam kisaran yang dapat diterima, menunjukkan bahwa model yang digunakan tidak memiliki multikolinearitas yang signifikan. |
| 14. | Regresi linear bivariat simpel dan aplikasinya pada data cuaca di Cilacap.[21] | Dataset dari BMKG Kabupaten Cilacap dari bulan Januari 2009 - Februari 2014. Dua variabel respon : curah hujan (Y1), kelembaban udara suatu wilayah (Y2) terhadap satu variabel respon : temperatur udara (X). | *Machine Learning* | Model dalam penelitian ini adalah menggunakan metode regresi linear bivariat simple. Model peramalan untuk Y1 adalah Y(1) = −894,130 + 45,892X yang berarti jika tempertur udara naik sebesar satu derajat Celcius maka curah hujan akan naik sebesar 45,892mm serta nilai rata-rata error bulanan = -0,00697mm dan MDEE(1) = 151,2132. Model peramalan untuk Y2 adalah Y(2) = 78,0433+0,1581X yang artinya jika tempertur udara naik sebesar satu derajat Celcius maka kelembaban udara akan naik 0,1581 persen seta nilai rata-rata error bulanan = 0,000441 persen dan MDE(2) = 1,206636. |
| 15. | Analisis regresi linier berganda dalam estimasi produktivitas tanaman padi di Kabupaten Karawang.[22] | Data Dinas Pertanian Kehutanan Perkebunan dan Peternakan Kabupaten Karawang. Variabel terikat : produktivitas padi (kw/ha). Data produktivitas padi dihasilkan dari data produksi dibagi luas panen. Variabel bebas : yaitu produksi, luas panen, luas tanam, rata-rata curah hujan, dan rata-rata hari hujan | *Machine Learning* | Model yang digunakan dalam penelitian ini adalah regresi liner berganda. Berdasarkan model regresi didapat 80,46% faktor-faktor produktivitas padi. Variabel-variabel yang mempengaruhi peningkatan jumlah produktivitas padi yaitu variabel produksi dan curah hujan, sedangkan variabel-variabel yang mempengaruhi penurunan jumlah produktivitas yaitu variabel luas panen, luas tanam, dan hari hujan. kesalahan relatif regresi yang diperoleh yaitu 4,642%. |
| 16. | Model regresi multivariat analisis kesejahteraan pedagang kaki lima berdasarkan karakteristik sosial ekonomi. [23] | Variabel bebas : Umur, Jenis kelamin, Status perkawinan, Tingkat pendidikan terakhir, Status migrasi, Tipe rumah, Sifat layanan pedagang kaki lima, Curahan jam kerja per hari, Lama menjalankan usaha, Jumlah tenaga kerja/karyawan diluar tenaga kerja keluarga yang membantu menjalankan usaha, Jumlah tenaga kerja keluarga yang membantu menjalankan usaha, Ketersediaan buku untuk pembukuan kegiatan usaha, Status Registrasi Usaha, Jenis dagangan, Sarana fisik pedagang kaki lima, Alternatif sumber pendapatan di luar pekerjaan sebagai PKL. Variabel dependen : Tingkat Kesejahteraan, dijabarkan dalam variabel: Pendapatan responden, Status Pekerjaan, dikelompokkan dalam: pekerja berusaha sendiri; berusaha sendiri dibantu oleh pekerja sementara/tidak dibayar; berusaha sendiri dibantu oleh pekerja permanen/dibayar; Karyawan/ Pekerja; Pekerja tidak dibayar. | *Machine Learning* | Model yang digunakan dalam penelitian adalah analisis *multivariate* yaitu analisis regresi linier dan *logistic ordinal*. Variabel yang signifikan berpengaruh terhadap rata-rata pendapatan pedagang kaki lima adalah tingkat pendidikan, curahan jam kerja, dan jumlah tenaga kerja diluar tenaga kerja keluarga yang ikut membantu. Model rata-rata pendapatan pedagang kaki lima adalah Y = -1.982 + 0.654pendidikan + 0.134curahan jam kerja + 0.817Jumlah tenaga kerja non keluarga. Variabel yang berpengaruh signifikan terhadap status pekerjaan pedagang kaki lima yaitu status perkawinan, sifat layanan dagangan, curahan jam kerja, serta jumlah tenaga kerja diluar tenaga kerja keluarga. Model dugaan yang menjelaskan status pekerjaan pedagang kaki lima yaitu G(x) = -16.308 - 0.519 status kawin + 0,739 sifat layanan + 1,19663 curahan jam kerja - 1,062 Jumlah tenaga kerja non keluarga. |