

Implementasi Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linear Berganda

Implementation of Data Mining Predictions for New Students Using Multiple Linear Regression Algorithm

Hendra Di Kesuma¹⁾, Deni Apriadi²⁾, Hengki Juliansa³⁾, Endang Etriyanti⁴⁾

¹Sistem Informasi, Ilmu Komputer, Universitas IGM Palembang

^{2,3,4}Sistem Informasi, STMIK Bina Nusantara Jaya Lubuk Linggau

Kota Palembang, Propinsi Sumatera Selatan

hendra.dikesuma@uigm.ac.id¹⁾, denidrv@gmail.com²⁾, hengki.juliansa@gmail.com³⁾,
endang.etriyanti@gmail.com⁴⁾

Abstract

The process of accepting new students is carried out every academic year with the selection stages that have been determined by the university in order to obtain the number of students and the quality of new students expected by the university. So far, STMIK Bina Nusantara Jaya Lubuklinggau doesn't yet have a system that can predict the number of new students in the coming period. . One solution that is needed is the application of data mining with multiple linear regression algorithms to be able to predict the number of new students in the coming year. The variables used consist of 3 variables, namely costs, registrants and new students where the variables are obtained based on the results of validity tests on many variables. Then the analysis was carried out with multiple linear regression algorithms, the prediction results of new students in 2023 were 38 people, then tested with a testing application using Ravidminer 5.0. where the prediction results of new students in 2023 are 38 people.

Keywords: Data Mining, Prediction, Multiple Linear Regression, New Student

Abstrak

Proses penerimaan mahasiswa baru dilakukan setiap tahun akademik dengan tahapan seleksi yang telah ditetapkan perguruan tinggi guna memperoleh jumlah mahasiswa dan kualitas mahasiswa baru yang diharapkan perguruan tinggi. Selama ini di STMIK Bina Nusantara Jaya Lubuklinggau belum memiliki sistem yang dapat memprediksi jumlah mahasiswa baru pada periode akan datang. Salah satu solusi yang diperlukan adalah penerapan data mining dengan algoritma regresi linear berganda untuk dapat memprediksi jumlah mahasiswa baru pada tahun yang akan datang. Variabel yang digunakan terdiri dari 3 variabel yaitu biaya, pendaftar dan mahasiswa baru dimana variabel diperoleh berdasarkan hasil uji validitas terhadap banyak variabel. Kemudian dilakukan analisa dengan algoritma regresi linear berganda diperoleh hasil prediksi mahasiswa baru pada tahun 2023 sebesar 38 orang, kemudian diuji dengan aplikasi pengujian menggunakan Ravidminer 5.0. dimana hasil prediksi mahasiswa baru pada tahun 2023 sebesar 38 orang.

Kata kunci: Data Mining, Prediksi, Regresi Linear Berganda, Mahasiswa Baru

1. Pendahuluan

Perguruan tinggi merupakan kelanjutan pendidikan pada jenjang yang lebih tinggi setelah menyelesaikan pendidikan menengah atas. Perguruan tinggi di Indonesia dapat berbentuk akademi, vokasi. Mempersiapkan para mahasiswa menjadi anggota masyarakat yang memiliki pengetahuan yang lebih luas dan dapat mengembangkan serta menyebarluaskan ilmu yang sudah diperoleh, merupakan tujuan dari perguruan tinggi. kemajuan suatu perguruan tinggi salah satunya ditentukan dari banyaknya mahasiswa baru yang masuk ke perguruan tinggi tersebut. Semakin meningkatnya jumlah mahasiswa baru, maka semakin baik keberadaan perguruan tinggi tersebut. STMIK Bina Nusantara Jaya Lubuklinggau terdiri dari 3 program studi yaitu Sistem Informasi, Manajemen Informatika dan Komputerisasi Akuntansi. Dalam menentukan jumlah mahasiswa baru ditahun berikutnya hanya melihat data mahasiswa baru yang ada disetiap tahunnya. Ada beberapa faktor

penyebab naik turunnya jumlah mahasiswa baru yaitu biaya, ketidakstabilan biaya kuliah menjadi salah satu penyebab naik turunnya mahasiswa baru, akreditasi prodi, memiliki akreditasi yang sangat baik dapat membuat banyak mahasiswa berminat untuk masuk keprodi tersebut. Jumlah pendaftar, calon mahasiswa diharuskan untuk melakukan pendaftaran terlebih dahulu sebelum melanjutkan ke tahap registrasi, hal ini dapat mempengaruhi jumlah mahasiswa baru yang masuk ke STMIK Bina Nusantara Jaya Lubuk Linggau, kemudian pandemi berdampak ke berbagai bidang, salah satunya yaitu bidang perekonomian, dikarenakan adanya pembatasan sosial dan karantina wilayah, sehingga banyak masyarakat yang kesulitan dalam mencari atau melakukan pekerjaan, hal ini dilakukan guna memenuhi biaya hidup sehari-hari termasuk didalamnya yaitu biaya pendidikan. Untuk memprediksi jumlah mahasiswa baru setiap program studi hanya melihat dari laporan kegiatan penerimaan mahasiswa baru yang dilaporkan pada PD. Dikti disetiap tahun nya.

Selama ini di STMIK Bina Nusantara Jaya Lubuk Linggau belum memiliki sistem yang dapat memprediksi jumlah mahasiswa baru. Untuk memprediksi jumlah mahasiswa baru hanya melihat dari laporan kegiatan penerimaan mahasiswa baru yang dilaporkan setiap tahunnya. Selama ini STMIK Bina Nusantara Jaya Lubuk Linggau belum memiliki sistem untuk memprediksi jumlah mahasiswa baru. Maka perlunya analisa dengan menerapkan konsep maining dalam melakukan prediksi mahasiswa baru menggunakan algoritma regresi linear berganda.

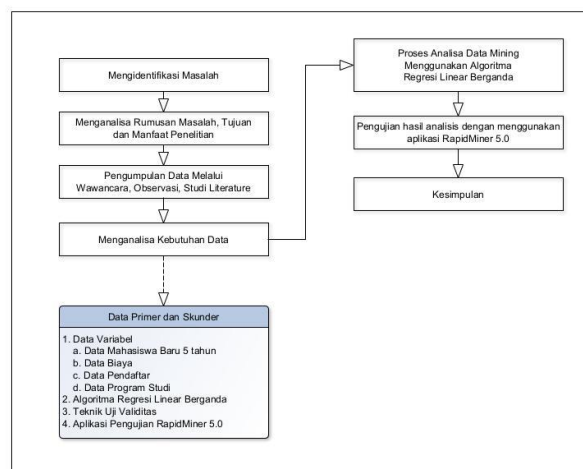
data mining ialah analisa atau pengamatan data dengan jumlah yang besar untuk menemukan hubungan yang belum diketahui sebelumnya, dan dua metode baru guna meringkas data supaya lebih mudah dimengerti serta kegunaannya untuk pemilih data[1],[2]. Analisis yang memiliki variabel bebas lebih dari satu disebut analisis regresi linier. Teknik regresi linier berganda digunakan untuk mengetahui ada tidaknya pengaruh signifikan antara dua atau lebih variabel bebas (X) terhadap variabel terikat atau tidak bebas (Y)[3],[4],[5]. dengan pemerapan data mining menggunakan algoritma regresi linier berganda dapat membantu dalam memprediksi jumlah mahasiswa baru yang akan di peroleh pada periode yang akan datang.

2. Tinjauan Pustaka

Berdasarkan analisis penelitian terdahulu menurut Nainggolan menjelaskan bahwa, permasalahan yang terdapat dalam penelitiannya di Sekolah Tinggi Agama Islam Raudhatul Akmal Batang Kuis yaitu, belum ada sistem data mining dalam mengestimasi jumlah mahasiswa baru, sehingga pihak kampus akan kesulitan dalam menyediakan fasilitas kampus, apabila jumlah mahasiswa baru yang mendaftara tiba-tiba bertambah, untuk mengatasi masalah tersebut maka perlu dibangun sistem data mining dengan menggunakan metode regresi linier berganda dengan variabel bebas yaitu promosi, jumlah lulusan dan fasilitas kampus[6]. Sedangkan menurut Boy, permasalahan yang terdapat dalam penelitiannya di Dinas Perkebunan Provinsi Sumatera Utara yaitu, belum ada sistem data mining untuk memprediksi harga minyak kelapa sawit, memprediksi harga minyak kelapa sawit sangat penting bagi instansi khususnya Dinas Perkebunan agar dapat mengambil kebijakan yang tepat dalam hal tata niaga minyak kelapa sawit, untuk mengatasi masalah tersebut maka perlu dibuat sistem data mining dengan menggunakan metode regresi linier berganda dengan variabel bebas yaitu produksi, harga dunia, ekspor dan konsumsi kosmetik[7]. Berdasarkan permasalahan diatas peneliti menarik kesimpulan bahwa, metode regresi linier berganda dapat digunakan untuk memprediksi mahasiswa baru, dengan menggunakan variabel terikat dan variabel bebas yang lebih dari satu. STMIK Bina Nusantara Jaya Untuk dapat membantu dalam mengetahui jumlah mahasiswa baru yang ada ditahun berikutnya

3. Metodologi Penelitian

Adapun metodologi penelitian yang digunakan pada penelitian ini dengan beberapa tahapan dan metode yang digunakan seperti pada gambar 1 sebagai berikut :



Gambar 1. Metodologi Penelitian

- a. Tahap pertama : dilakukan analisa permasalahan yang terjadi untuk mengidentifikasi masalah yang tepat dengan solusi penerapan data mining menggunakan algoritma regresi linear berganda dalam memprediksi mahasiswa baru pada periode yang akan datang.
- b. Tahapa Kedua : merumuskan masalah yang tepat sesuai dengan permasalahan yang telah diidentifikasi kemudian menentukan tujuan dan manfaat dari proses penelitian.
- c. Tahap Ketiga : melaukan pengumpulan data dengan menggunakan metode pengumpulan data dengan 3 teknik yaitu wawancara, observasi dan studiliteratur.
- d. Tahap Keempat : melakukan analisa kebutuhan data dengan menerapkan teknik pengujian yaitu uji validitas untuk memperoleh data yang dijadikan variabel yang tepat dalam melakukan prediksi mahasiswa baru untuk periode yang akan datang.
- e. Tahapa kelima : analisa dengan menggunakan algoritma regresi linear dengan menerapkan 3 variabel yang diperoleh dari hasil uji validitas dan menggunakan data histori mahasiswa baru 5 tahun terakhir.
- f. Tahapan keenam : pengujian menggunakan aplikasi pengujian RapidMiner 5.0 untuk mengetahui hasil prediksi.
- g. Tahapan ketujuh : hasil sistem prediksi mahasiswa baru menggunakan algoritma regresi linear berganda yang dapat memprediksi jumlah mahasiswa baru di setiap program studi di periode tahun akademik yang akan datang.

4. Hasil dan Pembahasan

Dari hasil analisa permasalahan dan rumusan masalah dan penetapan tujuan yang telah dilakukan. Maka dilakukan analisa kebutuhan data sehingga menghasilkan variabel dan data yang akan digunakan sebagai berikut :

4.1 Hasil Penentuan Variabel

Dalam penelitian ini, peneliti menggunakan variabel berdasarkan sumber refrensi dan hasil wawancara yang merujuk pada data hasil sebaran kuisioner, kemudian dilakukan pengolahan data kedalam bentuk kuantitatif, yaitu dengan cara menetapkan skor jawaban dari pernyataan yang telah dijawab oleh responden, dimana pemberian skor tersebut didasarkan pada ketentuan jumlah jawaban dari pernyataan yang telah dijawab oleh responden, :

Tabel 1. Alternatif Jawaban dan Bobot Nilai

No.	Alternatif	Bobot
1	Sangat Setuju	5
2	Setuju	4
3	Ragu-Ragu	3
4	Tidak Setuju	2
5	Sangat Tidak Setuju	1

Peneliti menggunakan 10 sample responden dalam uji validitas terhadap kuesioner, guna memilih variabel yang valid untuk digunakan sebagai variabel dalam prediksi mahasiswa baru, dalam penelitian ini uji validitas terhadap kuisioner menghasilkan nilai signifikasi dengan membandingkan nilai r hitung dengan r tabel untuk degree of freedom dengan alpha 5 % maka didapatkan r tabel= 0,6319. Hasil dari uji validitas dapat dilihat pada tabel dibawah ini [8],[9]:

Tabel 2. Perhitungan Uji Validitas

Responden	V ₁	V ₂	V ₃	V ₄	Y	Y ²
1	5	5	4	4	18	324
2	5	5	4	3	17	289
3	4	5	4	3	16	256
4	4	4	3	4	15	225
5	4	4	3	3	14	196
6	4	4	4	3	15	225
7	5	5	3	2	15	225
8	4	4	3	3	14	196
9	5	5	3	4	17	289
10	4	4	4	3	15	225
ΣX	44	45	35	32	156	2450
ΣX^2	196	205	125	106		
ΣXY	691	707	549	503		

Sehingga dihasilkan nilai Rhitung untuk V1 adalah 0,733213. Berdasarkan perhitungan yang telah dilakukan maka Rhitung untuk V1 adalah 0,733213, untuk mencari Rhitung V2,V3 dan V4, tahapan yang dilakukan sama dengan saat mencari Rhitung V1. Berikut ini merupakan tabel rekapitulasi hasil uji validitas :

Tabel 3. Hasil Uji Validitas

Butir	ΣX	ΣX^2	ΣXY	ΣY	ΣY^2	R hitung	R tabel	Hasil
V ₁	44	196	691	156	2450	0,733213	0,6319	Valid
V ₂	45	205	707	156	2450	0,780869	0,6319	Valid
V ₃	35	125	549	156	2450	0,468521	0,6319	Unvalid
V ₄	32	106	503	156	2450	0,49455	0,6319	Unvalid

4.2 Data Biaya dan Jumlah Pendaftar

Berdasarkan hasil uji validitas makan variabel yang digunakan pada penelitian ini yaitu, biaya dan jumlah pendaftar. Data yang diambil yaitu lima tahun terakhir dari tahun 2018-2022. Uraian hasil yang didapat bisa dilihat pada tabel 4 .

Tabel 4 Biaya dan Jumlah Pendaftar

Tahun	Biaya X ₁	Pendaftar X ₂	Mahasiswa Baru Y
2018	399	96	79
2019	399	42	30
2020	499	30	26
2021	499	26	24
2022	499	52	23

Setelah menentukan variabel selanjutnya memasukan nilai tersebut kedalam persamaan regresi linear berganda dapat di lihat pada tabel di bawah ini[10].

Tabel 5. Data Perhitungan Regresi

Biaya X ₁	Pendaftar X ₂	Mahasiswa Baru Y	X ₁ *Y	X ₂ *Y	X ₁ *X ₂	X ₁ ²	X ₂ ²
399	96	79	31521	7584	38304	159201	9216
399	42	30	11970	1260	16758	159201	1764
499	30	26	12974	780	14970	249001	900
499	26	24	11976	624	12974	249001	676
499	52	23	11477	1196	25948	249001	2704
2295	246	182	79918	11444	108954	1065405	15260

Pada tahap berikutnya menentukan nilai matrik dan determinan dimana nilai tersebut akan di gunakan untuk menentukan nilai konstanta a,b1 dan b2. Determinan

adalah nilai yang dapat di hitung dari unsur suatu matriks persegi. Adapun perhitungan sebagai berikut :
Menentukan Matrik A

$$N \quad \sum x_1 \quad \sum x_2$$

$$\sum x_1 \quad \sum x_1^2 \quad \sum x_1 x_2$$

$$\sum x_2 \quad \sum x_1 x_2 \quad \sum x_2^2$$

$$A = \begin{matrix} & 5 & 2295 & 246 \\ 2295 & 1065405 & 108954 \\ 246 & 108954 & 15260 \end{matrix}$$

$$H = \begin{matrix} \sum Y \\ \sum x_1 \cdot Y \\ \sum x_2 \cdot Y \end{matrix}$$

$$H = \begin{matrix} 182 \\ 79918 \\ 11444 \end{matrix}$$

Matrik A1, A2, A3

$$A1 = \begin{matrix} & 182 & 2295 & 246 \\ 79918 & 1065405 & 108954 \\ 11444 & 108954 & 15260 \end{matrix}$$

$$A2 = \begin{matrix} & 5 & 182 & 246 \\ 2295 & 79918 & 108954 \\ 246 & 11444 & 15260 \end{matrix}$$

$$A3 = \begin{matrix} & 5 & 2295 & 182 \\ 2295 & 1065405 & 79981 \\ 246 & 108954 & 11444 \end{matrix}$$

Kemudian diperoleh nilai determinan diperoleh dari nilai matrik yaitu :

Hasil Determinan A sebesar **111000000**, Hasil determinan A1 sebesar **3817956000**, determinan A2 sebesar **-7844000**, dan determinan A3 sebesar **77700000**
Kemudian menentukan nilai konstanta a, b1, dan b2 dengan hasil sebagai berikut :

$$a = \frac{\det(A1)}{\det(A)} = \frac{3817956000}{111000000} = 34,396$$

$$b1 = \frac{\det(A2)}{\det(A)} = \frac{-7844000}{111000000} = -0,070666667$$

$$b2 = \frac{\det(A3)}{\det(A)} = \frac{77700000}{111000000} = 0,7$$

Maka dapat dilakukan prediksi dengan persamaan sebagai berikut :

$$Y = a + b1.x1 + b2.X2$$

Jika data x1 (biaya) adalah 499, dan data x2(pendaftar) adalah 56 tahun 2022, maka Y untuk tahun 2023 adalah

$$Y = 34,396 + (-0,070666667).499 + 0,7.56$$

$$Y = \mathbf{38,608}$$

Berdasarkan perhitungan yang telah dilakukan pada data mahasiswa baru program studi sistem informasi dengan menggunakan persamaan algoritma regresi linier berganda didapatkan hasil prediksi 38 mahasiswa baru

untuk tahun 2023, Berikut tabel prediksi mahasiswa baru sampai dengan tahun 2027 seperti pada tabel

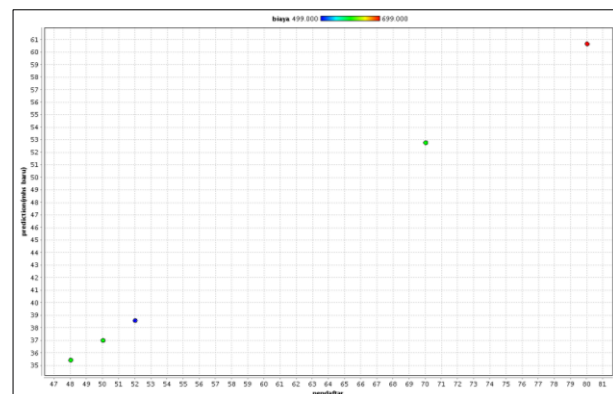
Tabel 6. Hasil Prediksi Tahun 2023 s/d 2027

No.	Tahun	Prediksi Maba
1	2023	38.608
2	2024	37.031
3	2025	35.454
4	2026	52.804
5	2027	60.690

Kemudian hasil analisa akan diuji menggunakan aplikasi bantu yaitu Rapidminer 5.0 [11]. Dimana Berdasarkan hasil pengujian analisa data menggunakan aplikasi rapidminer 5.0 diperoleh hasil prediksi sebagai berikut :

Row No.	sampel	mhs baru	prediction(...)	biaya	pendaftar
1	1.0	?	38.608	499	52
2	2.0	?	37.031	599	50
3	3.0	?	35.454	599	48
4	4.0	?	52.804	599	70
5	5.0	?	60.690	699	80

Gambar Data View Hasil Prediksi 2023-2027



Gambar Plot View Hasil Prediksi 2023-2027

Pada gambar 1 dan 2 menampilkan informasi hasil prediksi penerimaan mahasiswa baru mulai dari tahun 2023 sampai dengan 2027 memiliki hasil yang sama dengan analisa menggunakan algoritma regresi linear berganda.

5. Kesimpulan

Adapun kesimpulan yang diperoleh adalah sebagai berikut :

5.1 Simpulan

Berdasarkan dari uraian-uraian yang telah penulis jelaskan pada bab-bab sebelumnya pada Implementasi Data Mining Untuk Memprediksi Mahasiswa Baru

Dengan Algoritma Regresi Linear Berganda maka dapat di simpulkan yaitu:

1. Berdasarkan hasil dari uji validitas maka didapatkan variabel yang valid untuk digunakan sebagai variabel didalam penelitian ini, variabel tersebut adalah biaya dan jumlah pendaftar.
2. Berdasarkan hasil analisa menggunakan regresi linear berganda serta dengan aplikasi pengujian Rapidminer 5.0 menghasilkan hasil prediksi yang sama yaitu pada periode tahun 2023 diperoleh hasil prediksi mahasiswa baru pada program studi sistem informasi yaitu 38 orang.

5.2 Saran

Pada penelitian ini dilakukan hanya sebatas analisa dan pengujian menggunakan aplikasi rapidminer. Adapun saran untuk penelitian berikutnya agar pengembangan penelitian ini pada pembuatan sistem prediksi menggunakan bahasa pemrograman dengan menambahkan variabel yang lebih tepat sehingga hasil yang diperoleh dapat sesuai dengan yang diharapkan.

Ucapan Terima Kasih

Kami ucapkan terima kasih civitas Akademika Universitas UIGM Palembang dan STMIK Bina Nusantara Jaya Lubuklinggau yang telah mendukung sehingga penelitian ini dapat dilaksanakan dengan baik. Serta rekan sejawat yang telah membantu dalam penyelesaian penelitian ini.

Daftar Rujukan

- [1] Gaol, I. L. L., Sinurat, S. dan Siagian, E. R., 2019, Implementasi Data Mining dengan Metode Regresi Linier Berganda Untuk Memprediksi Data pada Persediann Buku Pada Pt Yudisthira Ghalia Indonesia Area Sumatera utara, Konferensi Nasional Teknologi Informasi dan komputer, 3, 1, 130- 133.
- [2] Adhiva, J., Mustakim, Putri, A. S. dan Setyorini, G. S., 2020, Prediksi Hasil Produksi Kelapa Sawit Menggunakan Regresi Pada PT. Perkebunan Nusantara V, Seminar Nasional Teknologi Informasi, Komunikasi dan Industri, 155-162.
- [3] Rivandi, A., Bu'ulolo, E. dan Silalahi, N., 2019, Penerapan Metode Regresi Linier Berganda Dalam Estimasi Biaya Pencetakan Spanduk (Studi Kasus: PT. Hansindo Setiaprata), Jurnal Pelita Informatika, 7, 3, 263-268.
- [4] Afkarina, N. K., Widodo, A. W. dan Furqon, M. T., 2019, Implementasi Regresi Linier Berganda Untuk Prediksi Jumlah Peminat Mata Kuliah Pilihan, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 3, 11, 10462-10467.
- [5] Gunawan, R., 2018, Implementasi Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Status Sosial Dan Kedisiplinan Pada Smk Bayu Pertiwi Menggunakan Metode Regresi Linier Berganda. Jurnal Saintikom (Jurnal Sains Manajemen Informatika Dan Komputer), 17(2), 175-183
- [6] Nainggolan, R. Z., Ibnutama, K., dan Suryanata, M. G., 2021, Implementasi Data Mining Dengan Metode Regresi Linier Berganda Dalam Estimasi Mahasiswa Baru Pada Sekolah Tinggi Agama Islam Raudhatul Akmal Batang Kuis. Jurnal Cyber Tech, 1(1).
- [7] Boy, A. F., 2020, Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar Domestik Menggunakan Algoritma Regresi Linier Berganda (Studi Kasus Dinas Perkebunan Provinsi Sumatera Utara). Journal of Science and Social Research, 3(2), 78-85.
- [8] Engkus, E., 2019, Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pasien Di Puskesmas Cibitung Kabupaten Sukabumi. Jurnal Governansi, 5(2), 99- 109.
- [9] Risandi, I. A., Muryati, M. dan Darma, B., 2021, Pengaruh Keselamatan Kerja dan Kesehatan Kerja Terhadap Kinerja Karyawan Bagian Panen Kelapa Sawit pada PT. Pratama Agro Sawit Terusan. JURNAL MAHASISWA, 1(1), 78-90.
- [10] Sulistyono, S. dan Sulistiyowati, W., 2018, Peramalan produksi dengan metode regresi linier berganda. PROZIMA (Productivity, Optimization and Manufacturing System Engineering), 1(2), 82-89
- [11] Yanto, R., 2020, Implementasi Data Mining Prediksi Kebutuhan Tenaga Listrik Di Kota Lubuklinggau. Techno. Com, 19(2), 197-206