

**PENERAPAN METODE REGRESI LINIER PADA ANALISIS
KETERSEDIAAN FORMASI JABATAN BERDASARKAN KUALIFIKASI
PENDIDIKAN PADA PENERIMAAN CASN DI LINGKUNGAN
PEMERINTAH KABUPATEN SE-PROVINSI JAWA TENGAH
DAN DIY MENGGUNAKAN PYTHON (TAHUN 2018-2021)**

SKRIPSI



disusun oleh

Awie Nuswatama

20.21.1487

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

**PENERAPAN METODE REGRESI LINEAR PADA ANALISIS
KETERSEDIAAN FORMASI JABATAN BERDASARKAN KUALIFIKASI
PENDIDIKAN PADA PENERIMAAN CASN DI LINGKUNGAN
PEMERINTAH KABUPATEN SE-PROVINSI JAWA TENGAH
DAN DIY MENGGUNAKAN PYTHON (TAHUN 2018-2021)**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Awie Nuswatama

20.21.1487

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

PERSETUJUAN

SKRIPSI

**PENERAPAN METODE REGRESI LINEAR PADA ANALISIS
KETERSEDIAAN FORMASI JABATAN BERDASARKAN KUALIFIKASI
PENDIDIKAN PADA PENERIMAAN CASN DI LINGKUNGAN
PEMERINTAH KABUPATEN SE-PROVINSI JAWA TENGAH
DAN DIY MENGGUNAKAN PYTHON (TAHUN 2018-2021)**

yang dipersiapkan dan disusun oleh

Awie Nuswatama

20.21.1487

telah disetujui oleh dosen Pembimbing Skripsi
pada tanggal 26 Desember 2022

Dosen Pembimbing,



Heri Sismoro, M.Kom.
NIK. 190302057

PENGESAHAN

SKRIPSI

**PENERAPAN METODE REGRESI LINEAR PADA ANALISIS
KETERSEDIAAN FORMASI JABATAN BERDASARKAN KUALIFIKASI
PENDIDIKAN PADA PENERIMAAN CASN DI LINGKUNGAN
PEMERINTAH KABUPATEN SE-PROVINSI JAWA TENGAH
DAN DIY MENGGUNAKAN PYTHON (TAHUN 2018-2021)**

yang dipersiapkan dan disusun oleh

Awie Nuswatama

20.21.1487

telah dipertahankan di depan Dewan Penguji
Pada tanggal 26 Desember 2022

Susunan Dewan Penguji

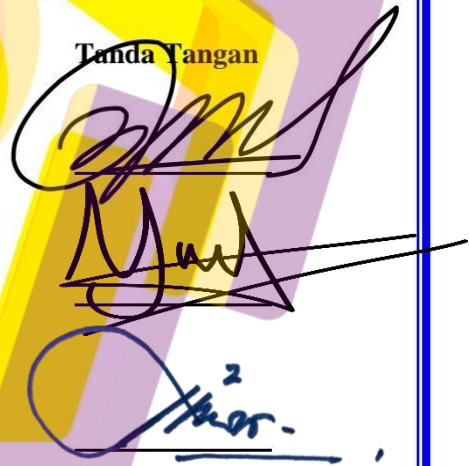
Nama Penguji

Ali Mustopa, M. Kom.
NIK. 190302192

Yoga Pristyanto, S.Kom, M.Eng.
NIK. 190302412

Heri Sismoro, M.Kom.
NIK. 190302057

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Desember 2022

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al-Fatta, M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Kebumen, 26 Desember 2022



Awie Nuswatama
NIM. 20.21.1487

HALAMAN MOTTO

“Dan sebaik-baik manusia adalah yang paling bermanfaat bagi manusia lainnya.”

~HR. Al-Qadlaa’iy dalam Musnad Asy-Syihaab no. 129, Ath-Thabaraaniy dalam Al-Ausath no. 5787~

“Hidup itu hanya berpindah dari masalah satu ke masalah lainnya.”

“Jangan jadi orang yang gampang ‘Gumunan’.”

~ Aji Kuncoro : 2013 ~



HALAMAN PERSEMBAHAN

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah serta karunia-Nya, sehingga penulis dapat menyelesaikan Skripsi ini dengan sebaik-baiknya dan mendapatkan hasil yang maksimal.

Tidak lepas dari bantuan berbagai pihak, oleh karena itu penulis ingin menyampaikan terima kasih kepada:

- 1 Allah SWT yang telah memberikan kelancaran dan segala kekuatan serta bantuan untuk dapat menyelesaikan penulisan ini.
- 2 Bapak dan Ibu, adik, dan juga semua yang selalu memberikan dukungan dan doa untuk kelancaran penulisan.
- 3 Bapak Heri Sismoro, M.Kom. selaku pembimbing, terima kasih atas bantuan dan bimbingan dalam penyelesaian penulisan ini.
- 4 Bapak dan Ibu Dosen Universitas Amikom Yogyakarta yang telah ikhlas dan sabar memberikan ilmu kepada mahasiswanya.
- 5 Seluruh Pegawai Dinas Komunikasi dan Informatika Kabupaten Kebumen atas dukungan dan doa untuk terselesaikannya penulisan ini.

KATA PENGANTAR

Segala Puji bagi Allah SWT, Rabb Semesta Alam. Shalawat dan salam semoga senantiasa ditujukan bagi Rasulullah SAW, keluarga, para sahabat, dan siapa saja yang meneladani mereka dengan baik hingga hari kiamat.

Dalam penyelesaian penulisan ini, banyak pihak yang telah memberikan bantuan baik secara moril, nasihat, dan semangat maupun materil. Atas segala bantuan yang telah diberikan, penulis ingin menyampaikan do'a dan ucapan terima kasih yang sedalam-dalamnya kepada semua pihak.

Penulis menyadari bahwa dalam penyusunan Skripsi ini masih banyak terdapat kekurangan. Oleh karena itu, penulis menerima saran dan kritik yang membangun dari pembaca sekalian. Penulis juga berharap agar Skripsi ini bisa memberikan manfaat kepada para pembaca khususnya, dan bagi penulis secara pribadi. *Aamiin ya Rabbal Alamin.*

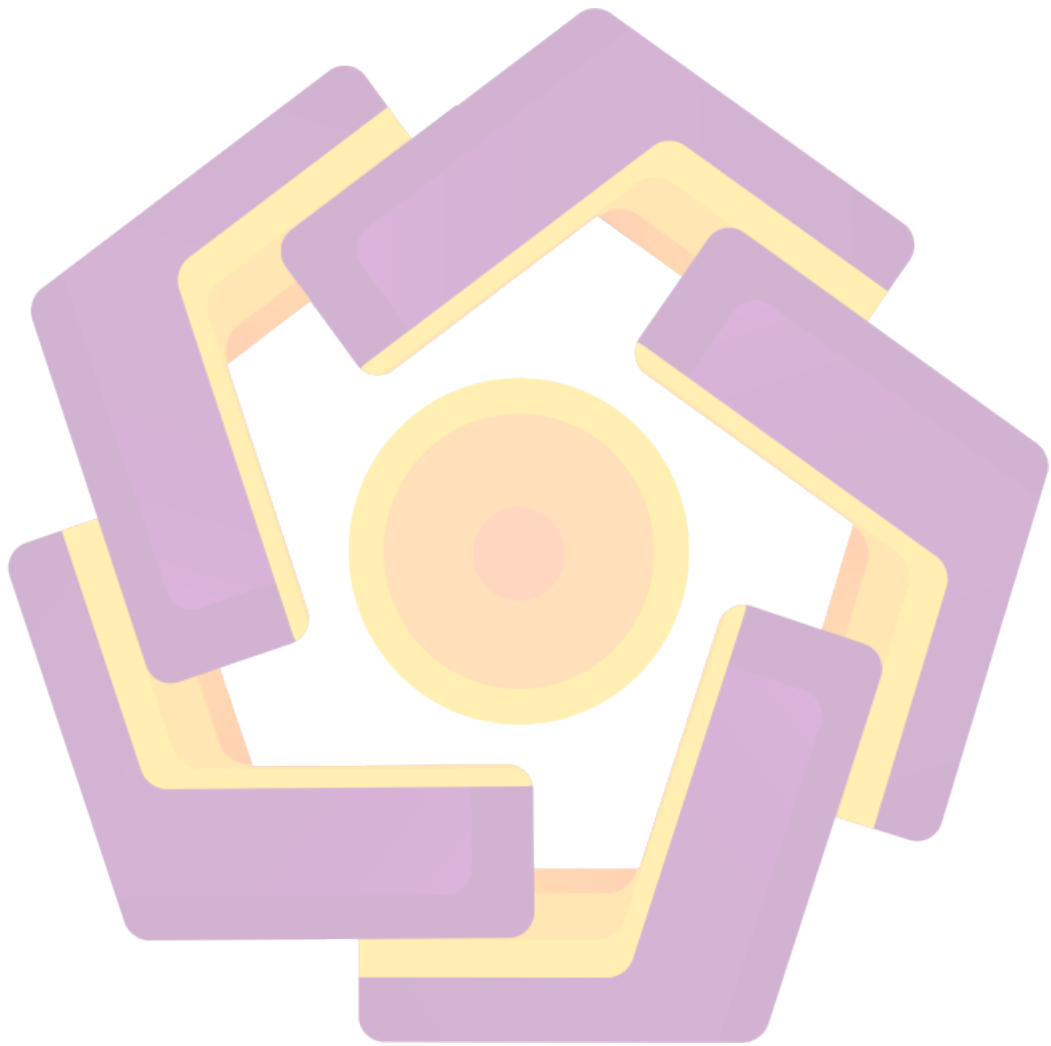
Kebumen, 26 Desember 2022
Penulis

Awie Nuswatama

DAFTAR ISI

SKRIPSI.....	i
PERSETUJUAN	ii
PENGESAHAN	iii
HALAMAN PERNYATAAN	iv
HALAMAN MOTTO	v
HALAMAN PERSEMBAHAN.....	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I.....	1
PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Maksud dan Tujuan Penelitian	4
1.5. Manfaat Penelitian.....	4
1.6. Metode Penelitian.....	5
1.7. Sistematika Penulisan.....	6
BAB II.....	7
LANDASAN TEORI.....	7
2.1 Kajian Pustaka	7
2.2 Data Mining.....	9
2.3 <i>Regression</i>	12
2.4 Metode Regresi Linear	13
2.5 Python.....	16
2.6 Pandas.....	16
2.7 Scikit-learn	17

2.8	NumPy (Numerical Python)	17
2.9	Matplotlib	17
2.10	Aparatur Sipil Negara (ASN)	17
BAB III		18
METODE PENELITIAN		18
3.1	Alur Penelitian	18
3.2	Alat dan Bahan	20
BAB IV		24
HASIL DAN PEMBAHASAN		24
4.1	Pengumpulan Data	24
4.1.1	Konversi File	24
4.1.2	Hasil Pengumpulan Data	24
4.1.3	<i>Datasets</i>	24
4.1.4	Data Kabupaten	25
4.2	<i>Data Preprocessing</i>	26
4.2.1	Data Cleaning	27
4.2.2	Data Transformation	27
4.3	Penentuan Data yang Dianalisis	28
4.4	Implementasi	29
4.4.1	Total Lowongan Sebagai Variabel Bebas	30
4.4.2	Total Lowongan Tenaga Kesehatan Sebagai Variabel Bebas	37
4.4.3	Menerapkan Aturan 80-20 (<i>Pareto Principle</i>) Pada Dataset	44
4.5	Matrik Hasil Implementasi	46
4.5.1	Total Kebutuhan sebagai Variabel X	46
4.5.2	Total Kebutuhan Tenaga Kesehatan sebagai Variabel X	46
BAB V		48
PENUTUP		48
5.1	Kesimpulan	48
5.2	Saran	49
DAFTAR PUSTAKA		xvi
LAMPIRAN		xviii



DAFTAR TABEL

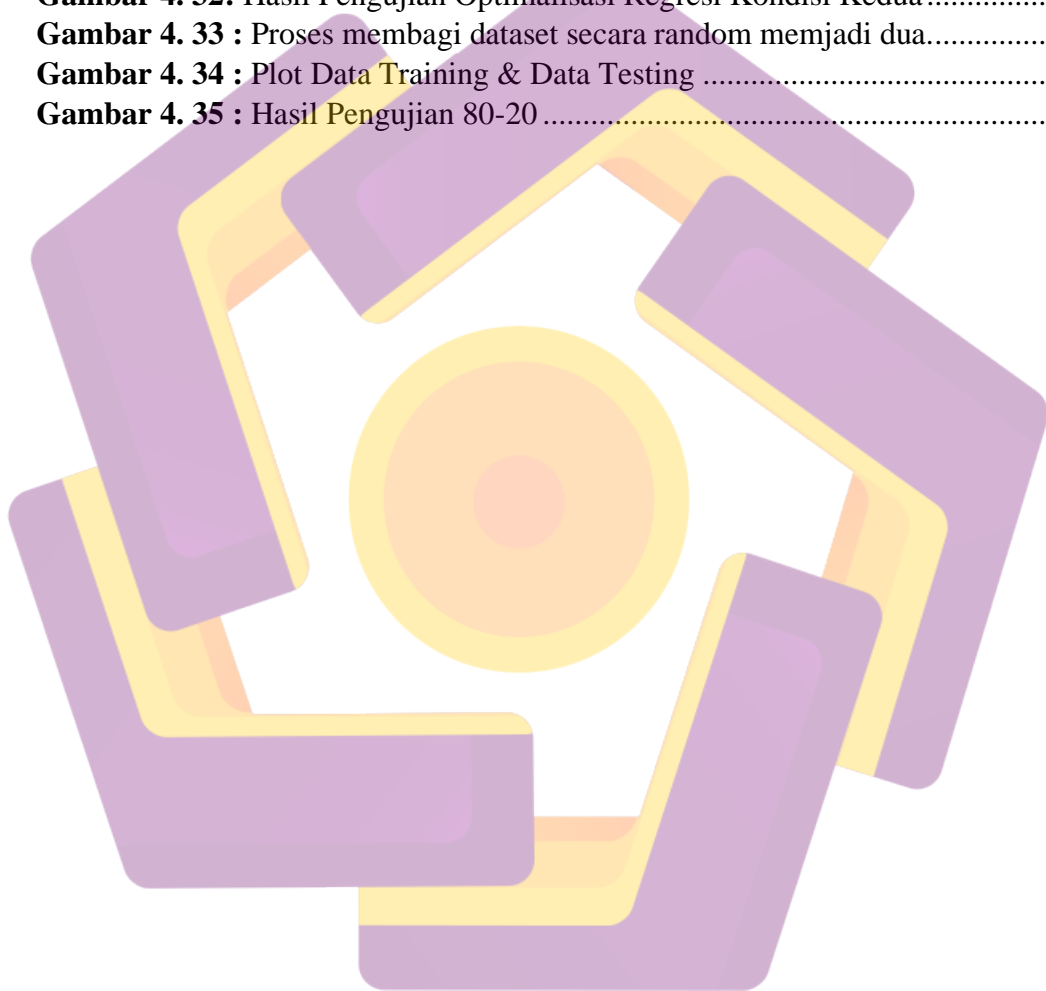
Tabel 2. 1 : Kajian Pustaka.....	8
Tabel 4.5. 1: Matrik Hasil Implementasi Pertama	46
Tabel 4.5. 2 : Matrik Hasil Implementasi Kedua	46



DAFTAR GAMBAR

Gambar 2. 1 : Rumus Regresi Linear	13
Gambar 2. 2: Rumus Menghitung a dan b pada Regresi Linear	14
Gambar 2. 3: Rumus Menghitung Error MAE.....	15
Gambar 2. 4: Rumus Menghitung Error MAPE	15
Gambar 2. 5: Rumus Menghitung Error MSE	15
Gambar 2. 6: Rumus Menghitung Error RMSE.....	15
Gambar 2. 7: Rumus Menghitung Error MAE%	16
Gambar 3. 1: Alur Penelitian.....	18
Gambar 3. 2: Versi Python	21
Gambar 3. 3: Versi Pandas	21
Gambar 3. 4: Versi Numpy	22
Gambar 3. 5: Versi Scikit-Learn	22
Gambar 3. 6: Versi Matplotlib	23
Gambar 4. 1: Import data .csv ke program.....	24
Gambar 4. 2: Dataset Utama	25
Gambar 4. 3: Data Kabupaten pada Dataset.....	26
Gambar 4. 4: Proses Cleaning Dataset	27
Gambar 4. 5: Mengubah Tipe data pada Kolom Dataset	28
Gambar 4. 6: Agregasi untuk memilih kualifikasi pendidikan	28
Gambar 4. 7: 10 Kualifikasi Pendidikan teratas.....	29
Gambar 4. 8: Menyiapkan Data untuk Regresi	30
Gambar 4. 9: Data Analisis Regresi Linear.....	31
Gambar 4. 10: Proses Perhitungan Regresi	32
Gambar 4. 11: Grafik Regresi Kondisi Pertama	33
Gambar 4. 12: Hasil Regresi Kondisi Pertama	33
Gambar 4. 13: Prediksi Regresi Kondisi Pertama	34
Gambar 4. 14: Hasil Prediksi Kondisi Pertama.....	34
Gambar 4. 15: Perintah untuk melakukan Pengujian Prediksi Regresi Kondisi Pertama.....	35
Gambar 4. 16: Hasil Pengujian Regresi Kondisi Pertama	35
Gambar 4. 17: Grafik Optimalisasi Regresi Kondisi Pertama	36
Gambar 4. 18: Hasil Optimalisasi Regresi Kondisi Pertama	36
Gambar 4. 19: Hasil Pengujian Optimalisasi Regresi Kondisi Pertama	37
Gambar 4. 20: Membuat dataframe tenaga kesehatan	37
Gambar 4. 21: Menyiapkan Data untuk Regresi Kondisi Kedua	38
Gambar 4. 22: Data Analisis Regresi Linear Kondisi Kedua.....	39
Gambar 4. 23: Proses Perhitungan Regresi Kondisi Kedua.....	39
Gambar 4. 24: Grafik Regresi Kondisi Kedua	40

Gambar 4. 25: Hasil Regresi Kondisi Kedua	40
Gambar 4. 26: Prediksi Regresi Kondisi Kedua	41
Gambar 4. 27: Data Prediksi Regresi Kondisi Kedua	41
Gambar 4. 28: Perintah untuk melakukan Pengujian Prediksi Regresi Kondisi Kedua	42
Gambar 4. 29: Hasil Pengujian Regresi kondisi Kedua	42
Gambar 4. 30: Grafik Optimalisasi Regresi Kondisi Kedua	43
Gambar 4. 31: Hasil Optimalisasi Regresi Kondisi Kedua	43
Gambar 4. 32: Hasil Pengujian Optimalisasi Regresi Kondisi Kedua	44
Gambar 4. 33 : Proses membagi dataset secara random menjadi dua.....	45
Gambar 4. 34 : Plot Data Training & Data Testing	45
Gambar 4. 35 : Hasil Pengujian 80-20	46



INTISARI

Di era industri 4.0 sekarang ini, internet dibanjiri dengan data yang begitu melimpah. Namun aset informasi tersebut masih memiliki ciri yaitu: *volume, velocity dan variety* yang sangat tinggi, sehingga diperlukan teknologi khusus untuk memprosesnya menjadi pengetahuan yang bernilai dalam proses pengambilan keputusan. Apabila melimpahnya aset informasi ini dimanfaatkan dengan baik, maka akan memberikan manfaat yang besar bagi suatu organisasi maupun individu, yaitu semakin bijaksana dalam pengambilan keputusan karena didasarkan pada data. Hal ini penting mengingat kondisi yang dapat berubah dengan begitu cepat. Tidak hanya keputusan dalam hal bisnis, namun bisa juga di berbagai bidang seperti di bidang pendidikan.

Salah satu *event* yang menghasilkan banyak data di internet yaitu proses seleksi penerimaan calon aparatur sipil negara(CASN). Data dari proses tersebut bisa dilakukan analisis, sehingga mampu memberikan pengetahuan yang bermanfaat, salah satunya untuk mempersiapkan proses seleksi pada penerimaan CASN periode berikutnya. Hal ini penting mengingat penerimaan CASN selalu mengundang antusiasme yang luar biasa dari para pelamar. Statistik dari data yang penulis himpun menunjukkan bahwa rata-rata untuk satu jabatan Calon Pegawai Negeri Sipil yang dibutuhkan, jumlah pendaftar rata-rata mencapai 14 orang.

Dengan membuat analisis terhadap data tersebut diatas, akan didapatkan berbagai informasi yang mampu menunjang keputusan. Baik dari sisi peserta, penyelenggara, maupun pemangku kebijakan di bidang pendidikan. Dengan menggunakan metode regresi linear informasi yang didapatkan berupa prediksi ketersediaan formasi jabatan berdasarkan kualifikasi pendidikan. Dari informasi tersebut diharapkan bisa dijadikan salah satu pertimbangan dalam memilih jurusan pendidikan. Bagi pemangku kebijakan di bidang pendidikan bisa menjadi pertimbangan dalam membuat keputusan mengenai ditambah atau dikurangnya jumlah mahasiswa yang masuk pada jurusan tertentu.

Kata Kunci : Data Mining, Regresi Linear, Python.

ABSTRACT

In the current industrial era 4.0, the internet is flooded with an abundance of data. However, these information assets still have characteristics, namely: very high volume, velocity and variety, so special technology is needed to process them into valuable knowledge in the decision-making process. If the abundance of information assets is put to good use, it will provide great benefits for an organization or individual, that is, they will be wiser in making decisions because they are based on data. This is important considering conditions can change so quickly. Not only decisions in terms of business, but also in various fields such as education.

One of the events that generates a lot of data on the internet is the process of selecting candidates for state civil servants (CASN). Data from this process can be analyzed, so as to be able to provide useful knowledge, one of which is to prepare for the selection process for the next CASN acceptance period. This is important considering that CASN acceptance always invites extraordinary enthusiasm from applicants. Statistics from the data that the authors have collected show that on average for one position the required Candidate for Civil Servants, the average number of applicants reaches 14 people.

By making an analysis of the data mentioned above, various information will be obtained that can support decisions. Both in terms of participants, organizers, and policy makers in the education sector. The information obtained can be in the form of the availability of position formations based on educational qualifications. This information can be used as a consideration in choosing a major in education. For policy makers in the field of education it can be a consideration in making decisions regarding adding or reducing the number of students enrolled in certain majors.

Keywords : Data Mining, Linear Regression, Python.