**USULAN**

**PENELITIAN INTERNAL**

**PEMODELAN BERBASIS DATA UNTUK MEMPREDIKSI GAJI BERDASARKAN FAKTOR-FAKTOR SPESIFIK DENGAN PENDEKATAN *MACHINE LEARNING***



Oleh:

|  |  |
| --- | --- |
| Syafrial Fachri Pane, ST., M.T.I., EBDP | 0416048803 |
| Amri Yanuar, ST., M.OT | 0412018603 |

**PROGRAM DIPLOMA IV TEKNIK INFORMATIKA**

**POLITEKNIK POS INDONESIA**

**TAHUN 2022**



**LEMBAR PENGESAHAN**

**Penelitian Internal**

|  |  |  |
| --- | --- | --- |
| Judul | : | Pemodelan berbasis data untuk memprediksi gaji berdasarkan faktor-faktor spesifik dengan pendekatan Machine Learning |
| **Ketua** |  |  |
| Nama Lengkap | : | Syafrial Fachri Pane, ST., M.T.I.,EBDP |
| Program Studi | : | D4 Teknik Informatika |
| NIDN | : | 0416048803 |
| Nomor HP | : | 08112164882 |
| Alamat Surel (e-mail) | : | [syafrial.fachri@poltekpos.ac.id](mailto:syafrial.fachri@poltekpos.ac.id) |
| **Pendamping1** |  |  |
| Nama Lengkap | : | Amri Yanuar, ST., M.MOT |
| NIDN | : | 0412018603 |
| Perguruan Tinggi | : | Politeknik Pos Indonesia |
| **Lama Penelitian** | : | **12 Bulan** |
| **Biaya Penelitian** | : | **Rp. 6.500.000 ,-** |

|  |  |  |
| --- | --- | --- |
| Mengetahui,  Ketua Program Studi |  | Bandung, 11-03-2022  Ketua |
|  |  |  |
| (Muhammad Yusril Helmi Setyawan, S.Kom, M.Kom.)  NIK/NIP. 113.74.163 |  | (Syafrial Fachri Pane, ST., M.T.I., EBDP)  NIK/NIP. 117.88.233 |
|  | Menyetujui,  Ketua LPM |  |
|  |  |  |
|  | (Sari Armiati, S.T., M.T.) NIK/NIP. 103.77.054 |  |

*11-03-2022 17:42:13 1/1 LPPM Politeknik Pos Indonesia*

# HALAMAN KETERLIBATAN MAHASISWA DALAM PENELITIAN

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nomor | Nama Mahasiswa | NPM | Bentuk Keterlibatan | Tanda Tangan |
| 1 | Bachtiar Ramadhan | 1204077 |  |  |
| 2 | Nur Tri Ramadhanti Adiningrum | 1204061 |  |  |

Bandung, …………………………………

Ketua Peneliti

Syafrial Fachri Pane, ST., M.T.I., EBDP

NIK. 117.88.233

# DAFTAR ISI

# RINGKASAN

Perkembangan ilmu pengetahuan dan teknologi pada Revolusi Industri 4.0 semakin berkembang pesat. Perubahan karakteristik pekerjaan adalah salah satu dampak tersendiri dari datangnya revolusi industri 4.0. Tentunya perusahaan perlu memiliki keunggulan manajemen yang efektif dalam menghadapi hal tersebut. Dengan demikian salah satu aspek yang berpengaruh besar terhadap kemajuan dan keberhasilan sebuah perusahaan adalah kinerja karyawannya. Oleh karena itu, penentuan gaji yang tepat oleh perusahaan adalah faktor internal terhadap kemajuan perusahaan. Sangat disayangkan, perkembangan perusahaan saat ini belum memliki suatu media keputusan untuk melakukan prediksi gaji karyawan berdasarkan kualitas data. Namun, untuk membuat keputusan bagaimana cara menentukan gaji karyawan dengan optimal perlu mempertimbangkan faktor-faktor lainya karena faktor tersebut merupakan bobot penilaian untuk mengukur kelayakan karyawan mendapatkan gaji.

Oleh karena itu, tujuan utama penelitian ini adalah membuat model prediksi gaji karyawan berdasarkan data. Karakteristik data yang digunakan terdiri dari umur, job level, total lama bekerja, masa bakti yang disebut faktor-faktor spesifik. Selanjutnya faktor-faktor spesifik akan diuji validitas dan korelasinya menggunakan pendekatan *machine learning* dengan metode *regression*. Tentunya hasil prediksi gaji karyawan perlu divisualisasikan secara realtime untuk dapat digunakan oleh perusahaan dalam menentukan keputusan dengan cepat. visualisasi hasil prediksi tersebut akan ditampilkan berbasis *web base* dengan framework Django. Rencana Keluaran penelitian ini adalah jurnal nasional terakreditasi SINTA 3, HAKI dan Buku ISBN yang diimplementasikan untuk referensi praktikum pada matakuliah Database di Prodi D4 Teknik Informatika Politeknik Pos Indonesia.

Kata Kunci : Pemodelan, Prediksi, Gaji, Regresi, faktor-faktor spesifik, *Machine Learning*

# BAB I PENDAHULUAN

## Latar Belakang

Perkembangan ilmu pengetahuan dan teknologi pada Revolusi Industri 4.0 semakin berkembang pesat. Revolusi Industri 4.0 sendiri mulai terjadi melalui rekayasa intelegensia dan internet of thing sebagai tulang punggung pergerakan dan konektivitas antara manusia dengan mesin[1]. Sehingga, terdapat penggabungan teknologi digital dan internet dengan industri konvensional, yang bertujuan untuk meningkatkan produktivitas, efisiensi dan layanan konsumen secara signifikan[2]. Era revolusi ini akan mendisrupsi berbagai kegiatan diberbagai bidang seperti pada bidang teknologi, ekonomi, sosial, dan politik[1]. Saat ini, kehidupan berada diawal revolusi yang secara mendasar mengubah cara hidup, bekerja, dan berhubungan satu sama lain [3].

Perubahan karakteristik pekerjaan adalah salah satu dampak tersendiri dari datangnya revolusi industri 4.0[4]. Karakteristik pekerjaan yang berubah akan mendisrupsi pekerjaan yang telah ada dan menggantikanya dengan pekerjaan dengan karakteristik baru [5]. Karakteristik baru pada pekerjaan juga membutuhkan kompetensi baru kepada para pekerja[6]. Tentunya perusahaan harus siap untuk saling bersaing dengan perusahaan yang lain[7]. Selanjutnya, perusahaan perlu memiliki keunggulan dan manajemen yang efektif untuk menghadapi persaingan tersebut[7]. Dengan demikian salah astu aspek yang berpengaruh besar terhadap kemajuan dan keberhasilan sebuah perusahaan adalah kinerja karyawannya[7]. Walaupun perusahaan tersebut memiliki teknologi yang canggih, namun tidak terdapat tenaga kerja didalamnya, perusahaan tidak akan dapat mencapai tujuannya[7].

Oleh karena itu, penentuan gaji yang tepat oleh perusahaan kepada karyawan adalah salah satu faktor yang berpengaruh secara internal terhadap kemajuan perusahaan. Selain itu, perusahaan juga harus bersedia mengeluarkan gaji bonus bagi karyawannya yang telah bekerja dengan maksimal dan sesuai dengan apa yang dibutuhkan oleh sebuah perusahaan. Sangat disayangkan, perkembangan perusahaan saat ini belum memliki suatu media keputusan untuk melakukan 2 prediksi gaji karyawan berdasarkan kualitas data.

Karakteristik dataset yang digunakan untuk memprediksi gaji karyawan terdiri dari parameter-parameter berdasarkan faktor-faktor spesifik. Selanjutnya faktor-faktor tersebut akan diuji validitas dan korelasinya menggunakan pendekatan machine learning. Faktor-faktor tersebut akan diambil berdasarkan pedoman interpretasi koefisien korelasi [8]. Untuk menentukan faktor yang dominan terhadap prediksi gaji, maka koefisien korelasi yang akan digunakan adalah tingkat hubungan sedang, kuat, dan sangat kuat. Metode yang digunakan pada machine learning yaitu regression. Regression digunakan untuk melakukan prediksi gaji karyawan. Tentunya hasil prediksi gaji karyawan perlu divisualisasikan secara realtime untuk dapat digunakan oleh perusahaan dalam menentukan keputusan dengan cepat. Visualisasi hasil prediksi tersebut akan ditampilkan berbasis web base dengan framework Django. Target luaran yang ingin dicapai pada penelitian ini adalah jurnal nasional terakreditasi SINTA 3, HAKI dan Buku ISBN yang diimplementasikan untuk referensi pembelajaran dan pratikum pada matakuliah Database di Program Studi Diploma 4 Teknik Informatika Politeknik Pos Indonesia.

# BAB II TINJAUAN PUSTAKA

## Tinjauan Studi

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

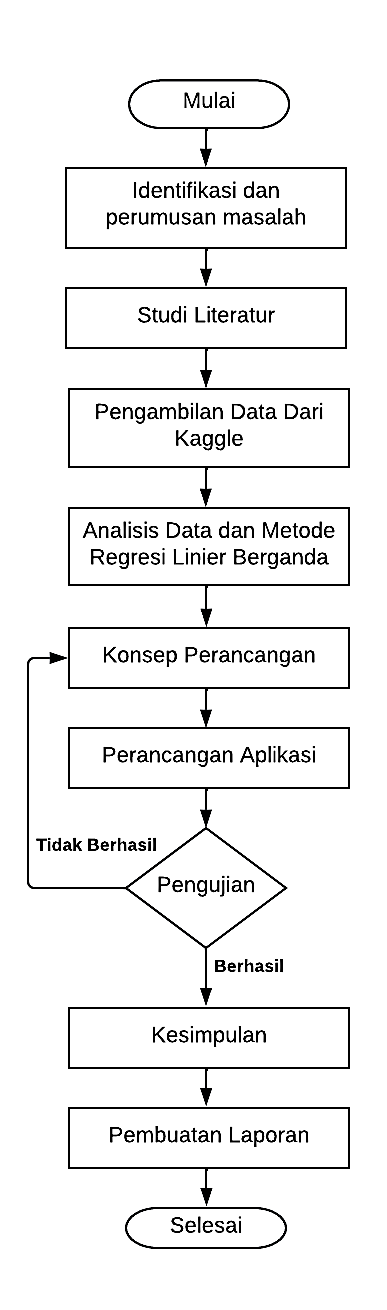
### Penelitian Terkait

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 variabel 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 dengan 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 model 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. |

# BAB III METODE PENELITIAN

## Diagram Alur Metodologi Penelitian

### Diagram Alur Metodologi Penelitian Penyelesaian Masalah

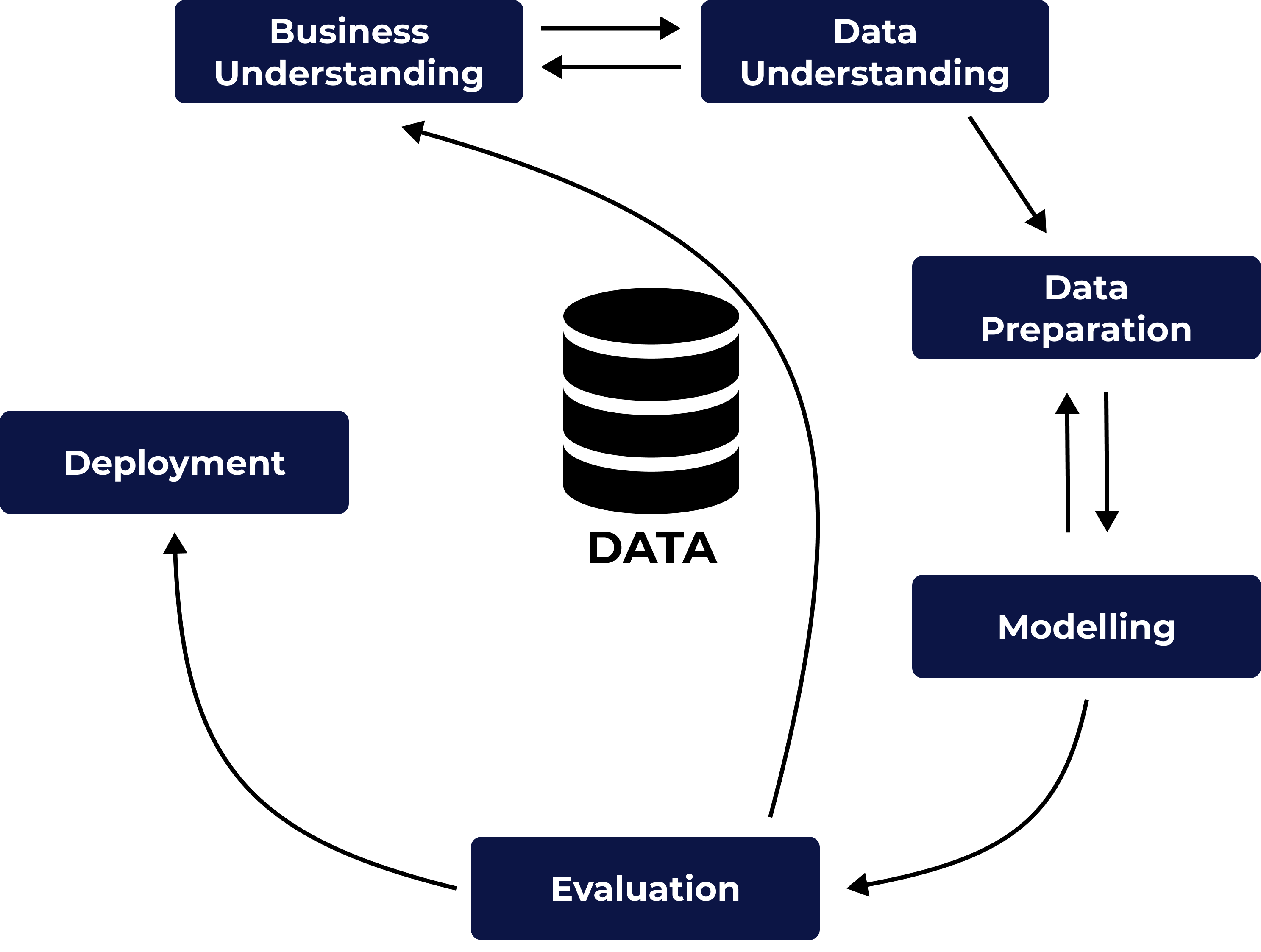


Berdasarkan diagram alur metodologi penelitian diatas, dapat dipaparkan sebagai berikut.

1. Hal pertama yang dilakukan untuk menyelesaikan masalah ialah melakukan pengidentifikasian masalah pada penelitian yang nantinya akan diselesaikan.
2. Setelah mengidentifikasi masalah yang ada, tahapan selanjutnya ialah melakukan studi literatur dengan mempelajari dan menemukan pengetahuan yang seluas-luasnya terhadap objek penelitian.
3. Tahap selanjutnya ialah mempersiapkan data dengan melakukan pencarian dan pengambilan data dari Kaggle.
4. Setelah didapatkannya data, data dapat dianalisis untuk kemudian diolah menggunakan model regresi linier berganda.
5. Setelah melakukan analisis dan pengolahan data hingga selesai, langkah selanjutnya ialah merancang konsep aplikasi yang akan dibuat.
6. Setelah konsep perancangan aplikasi terbuat, aplikasi dapat dirancang sesuai dengan konsep yang telah dibuat sebelumnya.
7. Ketika aplikasi telah dirancang, langkah selanjutnya ialah melakukan pengujian aplikasi. Apabila aplikasi berhasil dijalankan ketika melakukan pengujian, langkah selanjutnya dapat dilanjutkan. Namun apabila aplikasi tidak berhasil dijalankan ketika melakukan pengujian, maka akan kembali ke konsep perancangan untuk dilihat kembali apakah terdapat kesalahan atau tidak.
8. Jika aplikasi berhasil dijalankan sesuai konsep rancangan, maka dapat dibuatkan kesimpulan berdasarkan hasil rancangan.
9. Setelah itu dapat dibuatkan laporan berdasarkan hasil dari penelitian yang telah dilakukan.

### Diagram Alur Metodologi Pengolahan Data

Diagram alur metodologi pengolahan data yang digunakan ialah menggunakan metodologi CRISP-DM (Cross Industry Standard Process for Data Mining).



Berdasarkan diagram alur metodologi penelitian diatas, dapat dipaparkan sebagai berikut.

1. Business Understanding

Pada tahap ini merupakan tahap untuk menentukan seperti apa proses bisnis yang akan dibangun untuk melakukan hal apa yang ingin dicapai. Selain itu, dalam tahap ini juga diperlukan analisis apa saja alat dan bahan yang akan diperlukan untuk setiap fase proyek. Dalam hal ini, proses bisnis yang dilakukan pada penelitian adalah memprediksi gaji berdasarkan faktor-faktor spesifik. Prediksi gaji tersebut akan menggunakan data kepegawaian. Alat yang digunakan adalah berupa perangkat keras berupa laptop. Perangkat lunak yang digunakan ialah Jupyter Notebook, Anaconda, serta Microsoft Excel.

1. Data Understanding

Pada tahap data understanding, hal yang dilakukan adalah melakukan pemahaman bisnis yang berfokus untuk mengidentifikasi, mengumpulkan dan menganalisis kumpulan data untuk mencapai tujuan bisnis. Dalam hal ini, proses yang dilakukan ialah mengumpulkan data awal yang berasal dari Kaggle. Setelah data ditemukan, kemudian hal yang dilakukan adalah menjelajahi data dan melakukan verifikasi kualitas data (seberapa bersih atau otornya data yang digunakan).

1. Data Preparation

Tahap data preparation dilakukan untuk menyiapkan data akhir untuk dilakukannya pemodelan dengan cara memperbaiki masalah yang terdapat pada data. Hal yang dilakukan dalam penelitian ini adalah membersihkan data, serta memilih data yang akan digunakan dalam pemodelan. Sehingga nantinya akan didapatkan data akhir yang siap dimodelkan.

1. Modelling

Tahap modelling atau pemodelan adalah menentukan algoritma teknik pemodelan yang akan digunakan untuk menyelesaikan masalah yang ada. Dalam penelitian ini, model yang digunakan adalah menggunakan model regresi linier berganda. Karena model ini adalah model yang cocok untuk digunakan dalam melakukan prediksi terutama jika variabel yang mempengaruhi sangat banyak.

1. Evaluation

Tahap evaluation lebih berfokus untuk melihat apakah pemodelan yang dilakukan memiliki hasil yang baik atau malah sebaliknya. Pada tahap ini, evaluasi model dilakukan dengan menggunakan pengujian asumsi untuk melihat hasil-hasil evaluasi dari model.

1. Deployment

Yang terakhir ialah tahap deployment. Dimana pada tahap ini, hasil model penelitian yang telah dilakukan akan dikembangkan dan didokumentasikan. Pada tahap ini, hal yang dilakukan dalam penelitian yaitu mengembangkan model ke dalam bentuk aplikasi sehingga lebih mudah digunakan dalam melakukan proses prediksi. Selain itu, penelitian ini juga dibuatkan dalam bentuk laporan akhir untuk didokumentasikan.

# BAB IV BIAYA DAN JADWAL PENELITIAN

## Anggaran Biaya

Anggaran Penelitian mengacu pada komponen biaya Penelitian sebagai berikut :

|  |  |  |
| --- | --- | --- |
| No | Jenis Pengeluaran | Persentase |
| 1. | Gaji dan Upah |  |
| 2. | Bahan habis pakai |  |
| 3. | Perjalanan |  |
| 4. | Lain-lain (Publikasi, Seminar, Laporan, Lainnnya sebutkan) |  |

## Jadwal Penelitian

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Jenis Kegiatan | Bulan Ke- | | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | Kegiatan 1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |  |

# DAFTAR PUSTAKA

# LAMPIRAN-LAMPIRAN

## Lampiran 1. Format Jadwal Kegiatan

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Jenis Kegiatan | Bulan Ke- | | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | Kegiatan 1 |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |  |

## Lampiran 2. Susunan organisasi tim peneliti dan pembagian tugas

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Nama/NIDN** | **Asal Prodi** | **Bidang Ilmu** | **Alokasi Waktu (jam/minggu)** | **Uraian Tugas** |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |

## Lampiran 3. Justifikasi Anggaran

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. **Gaji dan Upah** | | | | |
| **Jabatan** | **Gaji & Upah / Jam (Rp)** | **Waktu (Jam/Minggu)** | **Minggu** | **TOTAL**  **(Rp)** |
| Syafrial Fachri Pane |  |  |  |  |
| Amri Yanuar |  |  |  |  |
| SUB TOTAL (Rp) | | | |  |
| 1. **Bahan Habis Pakai** | | | | |
| **Material** | **Justifikasi Pemakaian** | **Kuantitas** | **Harga Satuan**  **(Rp)** | **TOTAL (Rp)** |
| Material 1 |  |  |  |  |
| Material 2 |  |  |  |  |
| SUB TOTAL (Rp) | | | |  |
| 1. **Perjalanan** | | | | |
| **Material** | **Justifikasi Pemakaian** | **Kuantitas** | **Harga Satuan**  **(Rp)** | **TOTAL (Rp)** |
| Perjalanan ke tempat/kota - n | Survei/Sampling/dll |  |  |  |
| SUB TOTAL (Rp) | | | |  |
| 1. **Lain - Lain** | | | | |
| **Material** | **Justifikasi Pemakaian** | **Kuantitas** | **Harga Satuan**  **(Rp)** | **TOTAL (Rp)** |
| Lain-lain (administrasi, publikasi, seminar, laporan, lainnya sebutkan) | Publikasi/Laporan/dll |  |  |  |
| SUB TOTAL (Rp) | | | |  |
| TOTAL DANA YANG DIBUTUHKAN (Rp) | | | |  |

## Lampiran 4. Biodata ketua dan anggota tim pengusul

1. Identitas Diri

|  |  |  |
| --- | --- | --- |
| 1 | Nama Lengkap (dengan gelar) |  |
| 2 | Jenis Kelamin | L/P |
| 3 | Jabatan Fungsional Akademik |  |
| 4 | NIK |  |
| 5 | NIDN |  |
| 6 | Tempat dan tanggal lahir |  |
| 7 | E-mail |  |
| 8 | Nomor Telepon/ HP |  |
| 9 | Alamat Kantor |  |
| 10 | Nomor Telepon / Faks Kantor |  |
| 11 | Lulusan yang telah dihasilkan | D3 = ……. D4 = ……… |
| 12 | Mata Kuliah yang diampu | 1. |
| 2. |
| 3. |
| dst |

1. Riwayat Pendidikan

|  |  |  |  |
| --- | --- | --- | --- |
|  | S1 | S2 | S3 |
| Nama Perguruan Tinggi |  |  |  |
| Bidang Ilmu |  |  |  |
| Tahun Masuk/Tahun Lulus |  |  |  |
| Judul Skripsi/Tesis/Disertasi |  |  |  |
| Nama Pembimbing/Promotor |  |  |  |

1. Pengalaman Penelitian dalam 5 tahun Terakhir

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Tahun | Judul Penelitian | Pen | danaan |
| Sumber | Jml (Juta Rp) |
| 1. |  |  |  |  |
| 2. |  |  |  |  |

1. Pengalaman Pengabdian Kepada Masyarakat dalam 5 tahun terakhir

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Tahun | Judul Pengabdian Kepada Masyarakat | Pendanaan | |
| Sumber | Jml (Juta Rp) |
| 1. |  |  |  |  |
| 2. |  |  |  |  |

1. Publikasi Artikel Ilmiah Dalam Jurnal dalam 5 Tahun Terakhir

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Judul Artikel | Nama Jurnal | Volume/Nomor/Tahun |
| 1. |  |  |  |
| 2. |  |  |  |

1. Pemakalah Seminar Ilmiah (Oral Presentation) dalam 5 Tahun Terakhir

|  |  |  |  |
| --- | --- | --- | --- |
| No | Nama Pertemuan Ilmiah/ Seminar | Judul Artikel Ilmiah | Waktu dan Tempat |
| 1. |  |  |  |
| 2. |  |  |  |

1. Karya Buku/Buku Ajar/Praktek dalam 5 Tahun Terakhir

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Judul Buku | Tahun | Jumlah Halaman | Penerbit |
| 1. |  |  |  |  |
| 2. |  |  |  |  |

1. Perolehan HAKI dalam 5-20 Tahun Terakhir

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Judul/Tema HKI | Tahun | Jenis | Nomor P/ID |
| 1. |  |  |  |  |
| 2. |  |  |  |  |

1. Pengalaman Jabatan dalam 5 Tahun Terakhir

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Nama Jabatan | Tahun | Institusi | Tugas Pokok dan Fungsi |
| 1. |  |  |  |  |
| 2. |  |  |  |  |

1. Pengalaman Merumuskan Kebijakan Publik/Rekayasa Sosial Lainnya dalam 5 Tahun Terakhir

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Judul/Tema/Jenis Rekayasa  Sosial lainnya yang Telah Diterapkan | Tahun | Tempat Penerapan | Respon Masyarakat |
| 1. |  |  |  |  |
| 2. |  |  |  |  |

1. Penghargaan dalam 5 tahun terakhir (dari pemerintah, asosiasi atau institusi)

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Jenis Penghargaan | Institusi Pemberi Penghargaan | Tahun |
| 1. |  |  |  |
| 2. |  |  |  |

1. Organisasi/Profesi Ilmiah

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Tahun | Nama Organisasi/Profesi | Jenjang Jabatan/Keanggotaan |
| 1. |  |  |  |
| 2. |  |  |  |

Semua data yang saya isikan dan tercantum dalam porto folio ini adalah benar dan dapat dipertanggung jawabkan secara hukum. Demikian porto folio ini saya buat dengan sebenarnya.

Kota,Tangal-Bulan-Tahun

Pengusul,

**Tanda Tangan**

**(Nama Lengkap)**

## Lampiran 5. Surat Pernyataan Ketua Peneliti

**SURAT PERNYATAAN**

Yang bertanda tangan dibawah ini :

Nama : Syafrial Fachri Pane

NIDN : 0416048803

Pangkat / Golongan : ……………………………… Jabatan Fungsional : ………………………………

Dengan ini menyatakan bahwa proposal penelitian saya dengan judul  
**Pemodelan Berbasis Data Untuk Memprediksi Gaji Berdasarkan Faktor-Faktor Spesifik Dengan Pendekatan *Machine Learning*** yang diusulkan dalam skema penelitian Internal untuk tahun anggaran 20XX bersifat original dan belum pernah dibiayai oleh lembaga / sumber lain. Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku dan mengembalikan seluruh biaya penelitian yang sudah diterima ke kas Negara. Demikian pernyataan ini dibuat dengan sesungguhnya dan dengan sebenar-benarnya.

|  |  |
| --- | --- |
|  | Kota, Tanggal-Bulan-Tahun |
| Mengetahui,  Ketua Lembaga Penelitian    **Tanda Tangan**    (Sari Armiati, ST.,MT) NIP/NIK. | Yang Menyatakan,    **Tanda Tangan**    (Nama Lengkap)  NIP/NIK. |

## Lampiran 6. Format Catatan Harian (Logbook)

|  |  |  |
| --- | --- | --- |
| **No** | **Tanggal** | **Kegiatan** |
| 1 | …../……./….. | Catatan : …………………………………………………………………  …………………………………………………………………  …………………………………………………………………  Dokumen Pendukung :  1. Foto  2.Grafik/Tabel/ Dokumen  3.dst |
| 2 | …../……./….. | Catatan : …………………………………………………………………  …………………………………………………………………  …………………………………………………………………  Dokumen Pendukung :  1. Foto  2.Grafik/Tabel/ Dokumen  3.dst |
| 3 | …../……./….. | Catatan : …………………………………………………………………  …………………………………………………………………  …………………………………………………………………  Dokumen Pendukung :  1. Foto  2.Grafik/Tabel/ Dokumen  3.dst |
|  | dst |  |
|  |  |  |

## Lampiran 7. Format Surat Pernyataan Bebas Plagiat

**SURAT PERNYATAAN BEBAS PLAGIAT**

Saya yang bertanda tangan di bawah ini :

Nama :

NIDN :

Program Studi :

dengan ini menyatakan bahwa judul laporan ini **Pemodelan Berbasis Data Untuk Memprediksi Gaji Berdasarkan Faktor-Faktor Spesifik Dengan Pendekatan *Machine Learning*** benar bebas dari plagiat, dan apabila pernyataan ini terbukti tidak benar maka saya bersedia menerima sanksi sesuai ketentuan yang berlaku.

Demikian surat pernyataan ini saya buat untuk dipergunakan sebagaimana mestinya.

Bandung, …………………………..

Yang membuat pernyataan,

Materi 10.000

…………………………………….

## Lampiran 8. Format Surat Kuasa Ketua Peneliti

**Surat Kuasa Ketua Peneliti**

Saya sebagai ketua peneliti yang bertanda tangan di bawah ini :

Nama :

NIP :

Judul Penelitian :

Menguasakan kepada anggota saya,

Nama :

NIP :

untuk melaksanakan monitoring dan evaluasi program penelitian internal pada tanggal ……. bertempat di Politeknik Pos Indonesia.

Demikan surat ini saya buat untuk dipergunakan sebagaimana mestinya.

Bandung, …………………………..

Ketua Peneliti

Materai 10000

Nama