



# JEPIN

(Jurnal Edukasi dan Penelitian Informatika)

ISSN(e): 2548-9364 / ISSN(p) : 2460-0741

Vol. 6  
No. 3  
Desember  
2020

## Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode *Grid Search* pada Algoritma *Logistic Regression*

Muhamad Ichsan Gunawan<sup>#1</sup>, Dedy Sugiarto<sup>#2</sup>, Is Mardianto<sup>#3</sup>

<sup>#</sup>Jurusan Teknik Informatika Fakultas Teknologi Industri Universitas Trisakti  
Jl. Kyai Tapa No.1, RW.16, Grogol, Kec. Grogol petamburan, Jakarta Barat

<sup>1</sup>m.ichsan064001600023@std.trisakti.ac.id

<sup>2</sup>dedy@trisakti.ac.id

<sup>3</sup>mardianto@trisakti.ac.id

**Abstrak**— *Data Mining* tidak selalu digunakan untuk membahas tentang suatu bidang yang berkaitan dengan informatika, tetapi juga dapat digunakan untuk melakukan penelitian pada bidang lainnya, salah satunya bidang kesehatan untuk melakukan prediksi resiko gejala Diabetes Mellitus pada manusia dengan metode Regresi Logistik. Diabetes Mellitus merupakan salah satu penyakit terbanyak penderitanya di Indonesia. Menggunakan dataset yang berasal dari Pima Indians Diabetes Database dari model penelitian Lahiru Liyanapathirana, model tersebut memiliki tingkat akurasi 78%. Penelitian ini bertujuan untuk meningkatkan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode *Grid Search* pada Algoritma *Logistic Regression*. Sehingga didapat Model *Logistic Regression* dengan *Grid Search* pada *Classification Report* memiliki rata-rata akurasi model sekitar 79% dan akurasi data *check* sebesar 83,33%.

**Kata kunci**— *Classification Report, Data Mining, Diabetes Mellitus, Grid Search, Logistic Regression.*

### I. PENDAHULUAN

Diabetes Mellitus merupakan penyakit gangguan metabolik akibat Pankreas (organ di belakang perut) memproduksi sedikit insulin atau tidak ada insulin sama sekali. Insulin adalah hormon yang terbentuk secara alami, diproduksi oleh sel beta pankreas, yang membantu tubuh menggunakan gula untuk energi. Penderita penyakit Diabetes Mellitus pada tahun 2013 telah mencapai 382 juta [1], sehingga perlunya aksi nyata dalam mencegah penyakit Diabetes Mellitus.

Seiring dengan majunya teknologi informasi komunikasi utamanya pada bidang kecerdasan buatan, teknik pembelajaran mesin diperkenalkan untuk meningkatkan kemampuan sistem dalam membantu pendeteksian atau prediksi otomatis. Dengan bantuan sistem pembelajaran mesin, kemungkinan kesalahan diagnosis yang dilakukan oleh para ahli medis dapat dihindari, serta data medis dapat diperiksa dalam kurun

waktu yang singkat serta lebih detail [2]. Regresi logistik diharapkan dapat memberikan hasil yang sesuai dengan prediksi pada penelitian ini.

Penelitian yang dilakukan Lahiru Liyanapathirana tentang Alur Kerja Pembelajaran Mesin pada Data Diabetes Mellitus, pada penelitian tersebut membandingkan akurasi 7 model pembelajaran mesin terhadap data “Pima Indians Diabetes Database” dan didapat model Logistic Regression memiliki akurasi tertinggi sebesar 77% [3]. Kekurangan penelitian sebelumnya adalah akurasi model prediksi yang masih dibawah 80% sehingga perlunya peningkatan kinerja akurasi.

Diabetes Mellitus merupakan salah satu penyakit terbanyak penderitanya di Indonesia, untuk melakukan prediksi bisa saja terhambat karena bertambahnya data setiap harinya. Dengan berkembangnya teknologi internet saat ini maka berbagai sumber data yang berkaitan dengan data diabetes mellitus dapat mudah diperoleh secara real-time. Namun perkembangan varian dan volume data yang makin berkembang berakibat terhadap kebutuhan sistem komputasi yang sangat baik. Permasalahan ini dapat diatasi menggunakan Cloud Computing [4].

### II. TINJAUAN PUSTAKA

#### A. Diabetes Mellitus

Diabetes mellitus adalah sekelompok gangguan metabolisme, dengan gejalanya adalah kadar gula darah yang tinggi dengan rentan waktu yang lama pada penderitanya. Diabetes Mellitus adalah suatu kondisi di mana tubuh tidak memproduksi cukup hormon insulin, sehingga kadar gula dalam aliran darah tinggi. Ada banyak jenis diabetes; yang paling umum adalah diabetes tipe 1 dan tipe 2:

1) *Diabetes Tipe 1*: Ini hasil dari tidak membuat insulin. Penderita diabetes tipe 1 membutuhkan insulin, baik dengan injeksi atau menggunakan pompa insulin.

2) *Diabetes Tipe 2*: Ini hasil dari resistensi insulin, di mana sel-sel gagal menggunakan insulin dengan benar. Pasien diobati dengan kombinasi diet, olahraga, obat oral, atau kombinasi [5].

Terkadang beberapa penderita Diabetes Mellitus memiliki gejala yang tidak ada pada penderita Diabetes lain. Umumnya, orang dengan diabetes tipe 1 mengalami peningkatan rasa haus (polidipsia), sering buang air kecil (poliuria), dan meningkatnya rasa lapar (polifagia). Gejala dapat berkembang dari minggu ke bulan. Tidak diobati, kondisi ini dapat menyebabkan seseorang kehilangan kesadaran dan menjadi sangat sakit (ketoasidosis diabetik).

Orang dengan diabetes tipe 2 dan kehamilan mungkin memiliki gejala minimal. Pengukuran glukosa darah tunggal yang tinggi dapat membuat diagnosis, atau mungkin memerlukan beberapa pengukuran glukosa darah [2].

### B. Logistic Regression

Probabilitas adalah suatu peristiwa akan terjadi dari banyaknya peristiwa dalam suatu percobaan. Probabilitas selalu berkisar antara 0 dan 1. Odds didefinisikan sebagai probabilitas bahwa peristiwa akan terjadi dibagi dengan probabilitas bahwa peristiwa tidak akan terjadi.[6]

$$O(Y) = \frac{P(Y)}{1-P(Y)} \quad (1)$$

Regresi logistik adalah algoritma pembelajaran mesin yang paling terkenal setelah regresi linier. Dalam banyak hal, regresi linier dan regresi logistik serupa. Namun, perbedaan terbesar terletak pada apa yang mereka gunakan. Algoritma regresi linier digunakan untuk memprediksi / memperkirakan nilai tetapi regresi logistik digunakan untuk tugas klasifikasi [7].

Jenis pertanyaan yang bisa diperiksa oleh regresi logistik biner. Bagaimana kemungkinan terkena kanker paru-paru (ya vs tidak) berubah untuk setiap pon tambahan seseorang kelebihan berat badan dan untuk setiap bungkus rokok yang dihisap per hari? Apakah berat badan, asupan kalori, asupan lemak, dan usia memiliki pengaruh terhadap kemungkinan terkena serangan jantung (ya vs tidak)?

Regresi Logistik berusaha untuk mengidentifikasi apakah terdapat hubungan antara variabel dependen (Y) yang terkait dengan kejadian ya atau tidak dari suatu peristiwa (tipe dikotomis) dengan satu atau lebih variabel independen yang bertipe *categorical* atau *continuous*[8].

Di pusat analisis regresi logistik adalah tugas memperkirakan peluang log dari suatu peristiwa. Secara matematis, regresi logistik memperkirakan fungsi regresi linier berganda yang didefinisikan sebagai:

$$\begin{aligned} \ln\left(\frac{p}{q}\right) &= \ln(e^{\beta_0 + \beta_1 X}) \\ &= \beta_0 + \beta_1 X \end{aligned} \quad (2)$$

### C. Cloud Computing

Cloud computing adalah pengiriman sumber daya TI berdasarkan permintaan melalui Internet dengan harga bayar saat Anda bepergian. Alih-alih membeli, memiliki, dan memelihara pusat data dan server fisik, user dapat mengakses layanan teknologi, seperti daya komputasi, penyimpanan, dan basis data, sesuai kebutuhan dari penyedia cloud seperti Amazon Web Services (AWS).

1) *Infrastructure as a Service (IaaS)*: IaaS memberi pengguna kemampuan infrastruktur komputer dasar seperti penyimpanan data, server, dan perangkat keras - semuanya ada di cloud. IaaS memberi bisnis akses ke platform dan aplikasi besar tanpa perlu infrastruktur fisik di lokasi yang besar. Contoh : Amazon EC2, Azure VMs.

2) *Platform as a Service (PaaS)*: PaaS adalah lingkungan cloud yang mendukung pengembangan dan penyebaran aplikasi web. PaaS mendukung siklus hidup penuh aplikasi, membantu pengguna membangun, menguji, menyebarkan, mengelola, dan memperbarui semua di satu tempat. Contoh : Google App Engine.

3) *Software as a Service (SaaS)*: Model SaaS membuat perangkat lunak dapat diakses melalui aplikasi atau browser web. Beberapa program SaaS dapat diakses denggratis, tetapi banyak yang memerlukan berlangganan bulanan atau tahunan untuk mempertahankan layanan. Contoh : Dropbox dan Yahoo Mail [9].

### D. Grid Search

Parameter adalah variabel – variabel input pada suatu model. Sedangkan hyperparameter merupakan variabel yang dapat memengaruhi output model. Perbedaannya, nilai hyperparameter tidak diubah selama model dioptimisasi. artinya, nilai hyperparameter tidak bergantung pada data [10]. Grid Search adalah salah satu cara yang digunakan untuk menentukan kombinasi Hyperparameters [11].

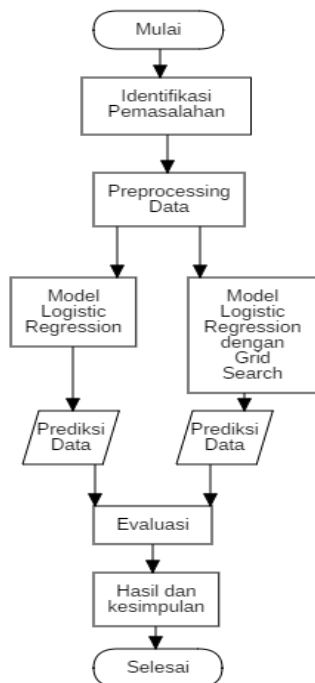
Metode *Grid Search* adalah alternatif untuk menemukan parameter terbaik untuk suatu model, sehingga pengklasifikasi dapat secara akurat memprediksi data yang tidak berlabel (Data pengujian). Metode ini dikategorikan sebagai metode lengkap untuk nilai parameter terbaik yang harus dieksplorasi masing-masing dengan menetapkan jenis nilai prediksi terlebih dahulu. Kemudian, metode akan menunjukkan skor untuk setiap nilai parameter untuk mempertimbangkan mana yang akan dipilih. Metode ini berlaku jika maksimum yang diperlukan diketahui berada dalam area terbatas yang ditentukan oleh batas atas dan bawah dari masing-masing variabel independen [12].

Prosedur *Grid Search* yang dilakukan Jenna Wong pada algoritma random forest, neural network, and support vector machine, mengungkapkan hasil penelitian bahwa terdapat nilai hyperparameter yang lebih baik daripada nilai *default SuperLearner package* [13]. Metrics pengukuran performa dari Grid Search adalah Mean Cross Validation (CV).

Grid Search membangun model untuk setiap kombinasi hyperparameter yang ditentukan dan mengevaluasi setiap

model. Teknik yang lebih efisien untuk penyetelan hyperparameter adalah pencarian acak - di mana kombinasi acak dari hiperparameter digunakan untuk menemukan solusi terbaik [12].

### III. METODOLOGI PENELITIAN



Gambar. 1 Tahapan penelitian

Peneliti mendapat ide identifikasi permasalahan dari media cetak nasional yang berisi tentang penyakit yang paling banyak dialami masyarakat Indonesia yaitu salah satunya Diabetes Mellitus. Perlu adanya pencegahan dan kesadaran masyarakat Indonesia terhadap penyakit Diabetes Mellitus. Penulis juga terinspirasi dari IBM Watson yang mengembangkan Sistem Artificial Intelligence pada bidang kesehatan yang memudahkan tenaga kerja medis untuk melakukan diagnosa alternatif, merekomendasikan tes, atau mengingatkannya tentang reaksi obat tertentu pada pasien.

Menggunakan dataset yang berasal dari Pima Indians Diabetes Database dari model penelitian Keshav Dhandhanian, model tersebut memiliki tingkat akurasi 78% [14], sehingga peneliti ingin melakukan peningkatan akurasi yang lebih tinggi dari model sebelumnya.

Melakukan analisa dari dataset Pima Indians Diabetes Database [15], dataset ini memiliki beberapa variabel, yaitu :

1. *Number of times pregnant* (Banyaknya kehamilan) - **preg**;
2. *Plasma glucose concentration* (Kadar glukosa) - **plas**;
3. *Diastolic blood pressure* (Tekanan darah) - **pres**;
4. *Triceps skin fold thickness* (Ketebalan kulit) - **skin**;
5. *2-h serum insulin* (Insulin) - **insu**;
6. *Body mass index* (Berat tubuh) - **mass**;

7. *Diabetes pedigree function* (Riwayat diabetes dalam keluarga) - **pedi**;
8. *Age* (Umur) - **age**;
9. *Class Variable* (Positif diabetes dan negative diabetes) - **class**.

```

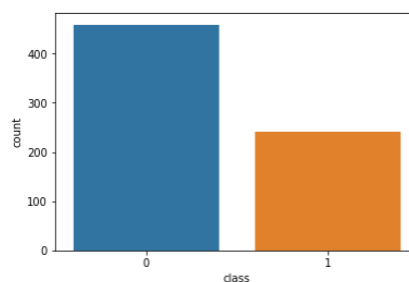
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
#   Column  Non-Null Count  Dtype
---  ---      -
0   preg    768 non-null    int64
1   plas    768 non-null    int64
2   pres    768 non-null    int64
3   skin    768 non-null    int64
4   insu    768 non-null    int64
5   mass    768 non-null    float64
6   pedi    768 non-null    float64
7   age     768 non-null    int64
8   class   768 non-null    int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
  
```

Gambar. 2 Data Info

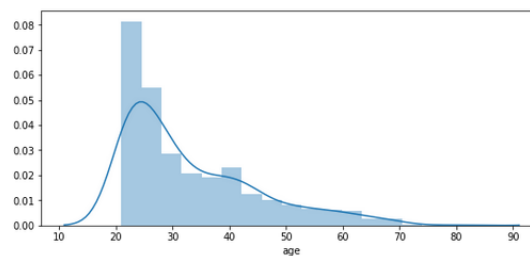
Pada informasi Gambar 2 terdapat 768 total data pasien, dari informasi yang didapat juga menunjukkan bahwa data sudah bersih (non-null) pada setiap kolomnya.

Melatih data dengan metode Logistic Regression pada Instance (*ml.t2.medium*) Jupyter Notebook yang dipasang pada Amazon Sagemaker sebagai penyedia *cloud computing*.

Pada dataset Pima Indians Diabetes Database Terdapat 768 total data pasien yang dibagi menjadi 300 data training, 400 data testing, dan 68 data checking. Pasien yang mengalami Diabetes (Class = 1) berjumlah sekitar setengah dari orang yang tidak terkena Diabetes (Class=0).

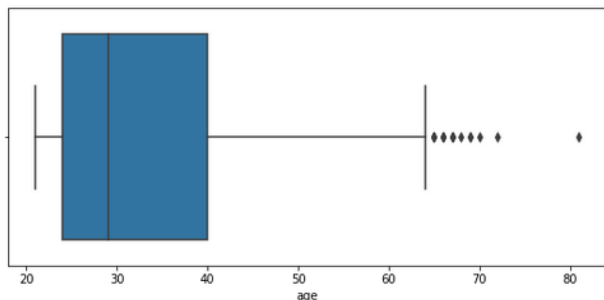


Gambar. 3 Grafik perbandingan jumlah class 0 dan 1



Gambar. 4 Grafik distplot variabel age

Dari distplot diatas dapat dilihat bahwa density dari data terletak di kisaran angka 21–30 tahun dan sangat jarang pasien yang berumur 20 tahun kebawah atau 80 tahun keatas.



Gambar. 5 bloxplot variabel age

Dari boxplot menunjukkan bahwa data memiliki outlier. Model Logistic Regression yang dibuat berasal dari 8 Variabel Independen yaitu *preg*, *plas*, *pres*, *skin*, *insu*, *mass*, *pedi*, dan *age*. Serta Variable Dependen *class*. Dari model tersebut, dihasilkan *classification report* pada gambar. 4

```
print(classification_report(y_test, y_pred)) #Check performa
```

	precision	recall	f1-score	support
0	0.80	0.89	0.84	262
1	0.71	0.56	0.63	131
micro avg	0.78	0.78	0.78	393
macro avg	0.76	0.73	0.74	393
weighted avg	0.77	0.78	0.77	393

Gambar. 6 Classification report

Dari classification report dapat dilihat bahwa model memiliki rata-rata akurasi sekitar 77% mulai dari precision, recall, f1-score, dan support. Accuracy juga menunjukkan nilai sebesar 78%.

```
roc_auc_score(y_test, y_pred)
```

0.7251908396946566

Gambar. 7 AUC score

Dari Hasil AUC Score didapatkan score sebesar 72.5%. AUC singkatan dari "Area under the ROC Curve" artinya, AUC mengukur seluruh area dua dimensi di bawah seluruh kurva ROC. AUC membuat mudah dalam membandingkan model satu dengan yang lainnya dengan memilih model yang AUC-nya paling besar[16].

Melakukan *Tuning Hyperparameters Logistic Regression* menggunakan *Grid Search*. *List Hyperparameters* yang akan diuji [11]:

- *penalty* = ['l1', 'l2']
- *C* = np.logspace(-2,2,10)

Dari parameter *penalty* dan *c* didapat Nilai *Hyperparameters* Terbaik yaitu:

- *Best Penalty*: l1
- *Best C*: 4.6415888336127775.

Hasil *Hyper Parameter* model, dihasilkan *classification report* pada gambar. 6

```
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.82	0.89	0.85	262
1	0.73	0.62	0.67	131
micro avg	0.80	0.80	0.80	393
macro avg	0.78	0.75	0.76	393
weighted avg	0.79	0.80	0.79	393

Gambar. 8 Classification report hasil hyper parameter

Dari classification report dapat dilihat bahwa model memiliki rata-rata akurasi sekitar 79% mulai dari precision, recall, f1-score, dan support. Accuracy juga menunjukkan nilai sebesar 80%.

```
roc_auc_score(y_test, y_pred)
```

0.7519083969465649

Gambar. 9 AUC score hasil hyper parameter

Dari Hasil AUC Score didapatkan score sebesar 75.1%

#### IV. HASIL DAN PEMBAHASAN

Dari model Logistic Regression hasil Hyper Parameter dengan Grid Search dilakukan pengecekan menggunakan 18 data check terakhir dari nomer 751 sampai 768 dengan nilai Variabel Independen yang berbeda – beda.

TABEL I  
DATA PIMA INDIANS DIABETES DATABASE NO 751 - 768

no	preg	plas	pres	skin	insu	mass	pedi	age
751	4	136	70	0	0	31.2	1.182	22
752	1	121	78	39	74	39	0.261	28
753	3	108	62	24	0	26	0.223	25
754	0	181	88	44	510	43.3	0.222	26
755	8	154	78	32	0	32.4	0.443	45
756	1	128	88	39	110	36.5	1.057	37
757	7	137	90	41	0	32	0.391	39
758	0	123	72	0	0	36.3	0.258	52
759	1	106	76	0	0	37.5	0.197	26
760	6	190	92	0	0	35.5	0.278	66
761	2	88	58	26	16	28.4	0.766	22
762	9	170	74	31	0	44	0.403	43
763	9	89	62	0	0	22.5	0.142	33
764	10	101	76	48	180	32.9	0.171	63
765	2	122	70	27	0	36.8	0.34	27
766	5	121	72	23	112	26.2	0.245	30
767	1	126	60	0	0	30.1	0.349	47
768	1	93	70	31	0	30.4	0.315	23

Tabel I tersebut sudah memiliki variabel dependen yaitu "class" yang sudah dipisahkan di Tabel II. Nilai "class" akan dicek dengan nilai prediksinya model *Logistic Regression* tanpa *Grid Search* dan model *Logistic Regression* dengan *Grid Search*. Hasil perbandingannya Tabel II.

TABEL II  
PERBANDINGAN PREDIKSI VARIABEL CLASS

no	class	Hasil Prediksi LR	
		Tanpa Grid Search	Dengan Grid Search
751	1	0	1
752	0	0	0
753	0	0	0
754	1	1	1
755	1	1	1
756	1	0	0
757	0	1	0
758	1	0	0
759	0	0	0
760	1	1	1
761	0	0	0
762	1	1	1
763	0	0	0
764	0	0	0
765	0	0	0
766	0	0	0
767	1	0	0
768	0	0	0
	benar	13	15
	salah	5	3
	akurasi	72,22%	83,33%

Hasilnya Prediksi *Logistic Regression* tanpa Grid Search yaitu 13 data benar dan 5 data salah dari 18 data cek, dengan akurasi 72,22%. Sedangkan Prediksi *Logistic Regression* dengan Grid Search yaitu 15 data benar dan 3 data salah dari 18 data cek, dengan akurasi 83,33%. Sehingga terjadi peningkatan akurasi sebesar 11,11%.

V. KESIMPULAN

Hasil penelitian peningkatan kinerja akurasi prediksi penyakit diabetes mellitus terhadap perbandingan model Logistic Regression tanpa Grid Search dan dengan Grid Search terjadi peningkatan yang signifikan. Model Logistic Regression dengan Grid Search pada *Classification Report* memiliki rata-rata akurasi sekitar 79% mulai dari *precision*, *recall*, *f1-score*, dan *support*. *Accuracy* juga menunjukkan nilai sebesar 80%.

Grid Search menentukan kombinasi Hyperparameters terbaik dengan teknik pengukuran Mean Cross Validation (CV) yang mengkombinasikan nilai yang dimasukkan pada Hyperparameters untuk melihat kombinasi nilai yang memiliki *CV Score* tertinggi.

Dari Hasil AUC Score Model *Logistic Regression* dengan *Grid Search* didapatkan score sebesar 75.1%.

Pengecekan pada data cek didapatkan model Prediksi *Logistic Regression* dengan *Grid Search* yaitu 15 data benar dan 3 data salah dari 18 data cek, dengan akurasi sebesar 83,33%.

Dari pemodelan Prediksi *Logistic Regression* dengan *Grid Search* penelitian yang dilakukan ini, dapat digunakan

dalam pembuatan aplikasi berbasis web untuk mendeteksi penyakit diabetes mellitus.

REFERENSI

[1] M. F. Rahman, M. I. Darmawidjadja, and D. Alamsah, "KLASIFIKASI UNTUK DIAGNOSA DIABETES MENGGUNAKAN METODE BAYESIAN REGULARIZATION NEURAL NETWORK ( RBNN )," *J. Inform.*, vol. 11, no. 1, pp. 36–45, 2017.

[2] Noviandi, "IMPLEMENTASI ALGORITMA DECISION TREE C4 . 5 UNTUK PREDIKSI PENYAKIT DIABETES," *J. INOHIM*, vol. 6, no. 1, pp. 1–5, 2018.

[3] L. Liyanapathirana, "Machine Learning Workflow on Diabetes Data: Part 01," *towards data science*, 2018. [Online]. Available: <https://towardsdatascience.com/machine-learning-workflow-on-diabetes-data-part-01-573864fcc6b8>. [Accessed: 25-Sep-2019].

[4] J. Tandy and Siswono, "CLOUD COMPUTING DAN DAMPAKNYA TERHADAP BISNIS," *ComTech*, vol. 4, no. 9, pp. 687–695, 2013.

[5] B. He, K. Shu, and H. Zhang, "Machine Learning and Data Mining in Diabetes Diagnosis and Treatment," *IOP Conf. Ser. Mater. Sci. Eng. Pap.*, vol. 490, pp. 1–6, 2019.

[6] J. Harlan, *ANALISIS REGRESI LOGISTIK*. Penerbit Gunadarma, 2018.

[7] W. Widhiarso, "Berkenalan dengan Regresi Logistik," *UGM*, 2012. [Online]. Available: <https://simpan.ugm.ac.id/s/RTwA1Ti5b0F8vkr#pdfviewer>. [Accessed: 21-Nov-2019].

[8] I. Unggara, A. Musdholifah, and A. K. Sari, "Optimization of ARIMA Forecasting Model Using Firefly Algorithm," *Indones. J. Comput. Cybern. Syst.*, vol. 13, no. 2, pp. 127–136, 2019.

[9] R. I. A. BUDIASTUTI, "PREDIKSI INDEKS HARGA KONSUMEN HARIAN MENGGUNAKAN SUPPORT VECTOR REGRESSION BERBASIS CLOUD COMPUTING," INSTITUT TEKNOLOGI SEPULUH NOPEMBER SURABAYA, 2017.

[10] P. Probst, A.-L. Boulesteix, and B. Bischl, "Tunability : Importance of Hyperparameters of Machine Learning Algorithms," *J. Mach. Learn. Res.*, vol. 20, pp. 1–32, 2019.

[11] A. Martulandi, "Tuning Hyperparameters Logistic Regression Menggunakan Grid Search," *medium.com*, 2019. [Online]. Available: <https://medium.com/@adiptamartulandi/tuning-hyperparameters-logistic-regression-menggunakan-grid-search-ucupstory-fb1ab9db082a>. [Accessed: 15-Mar-2020].

[12] M. M. RAMADHAN, I. S. SITANGGANG, F. R. NASUTION, and A. GHIFAR, "Parameter Tuning in Random Forest Based on Grid Search Method for Gender Classification Based on Voice Frequency," *Int. Conf. Comput. Electron. Commun. Eng.*, pp. 625–629, 2017.

[13] J. Wong, T. Manderson, M. Abrahamowicz, D. L. Buckeridge, and R. Tamblyn, "Can Hyperparameter Tuning Improve the Performance of a Super Learner ? A Case Study," *Epidemiology*, vol. 30, no. 4, pp. 521–531, 2019.

[14] K. Dhandhanian, "End-to-End Data Science Example: Predicting Diabetes with Logistic Regression," *Towards Data Science*, 2018. [Online]. Available: <https://towardsdatascience.com/end-to-end-data-science-example-predicting-diabetes-with-logistic-regression-db9bc88b4d16>. [Accessed: 10-Jan-2020].

[15] M. Barale and D. Shirke, "Cascaded Modeling for PIMA Indian Diabetes Data," *Int. J. Comput. Appl.*, vol. 139, pp. 1–4, 2017.

[16] T. T. Maskoen and D. Purnama, "Area Under the Curve dan Akurasi Cystatin C untuk Diagnosis Acute Kidney Injury pada Pasien Politrauma," *Maj. Kedokt. Bandung*, vol. 50, no. 4, pp. 259–264, 2018.