

Optimasi Model Prediksi Kesuksesan Startup Menggunakan StandartScaler Transform

Wulan Sri Lestari¹

¹Teknologi Informasi, Fakultas Informatika, Universitas Mikroskil Medan

E-mail: *¹wulan.lestari@mikroskil.ac.id

Abstrak – Pertumbuhan pesat startup di era teknologi modern menimbulkan tantangan kompleks dalam memprediksi kesuksesan suatu startup. Berdasarkan data yang diperoleh, 90 dari 100 startup gagal mencapai tingkat pendanaan yang diinginkan. Keberhasilan sebuah startup tidak hanya mencerminkan prestasi bisnis individu tetapi juga memiliki dampak yang luas pada dinamika ekonomi global. Oleh karena itu, perlu dikembangkan model prediksi yang handal sebagai dasar pengambilan keputusan yang efektif dalam pengembangan startup. Untuk mengatasi masalah ini, penelitian ini menggunakan integrasi Deep Neural Networks dengan StandartScaler Transform untuk mengoptimalkan model prediksi kesuksesan startup. StandartScaler Transform digunakan untuk menormalkan distribusi data sebelum diterapkan pada Deep Neural Networks. Hasil eksperimen pada model prediksi kesuksesan startup menunjukkan peningkatan akurasi sebesar 35,25% pada tahap training, dan peningkatan tambahan sebesar 36,95% pada tahap testing. Model prediksi kesuksesan startup yang dibangun berhasil mengatasi masalah overfitting/underfitting, sehingga menjadikannya sebagai dasar untuk pengembangan aplikasi prediksi kesuksesan startup.

Kata Kunci — Kesuksesan Startup, Deep Neural Networks, StandartScaler Transform, Optimasi, Prediksi

1. PENDAHULUAN

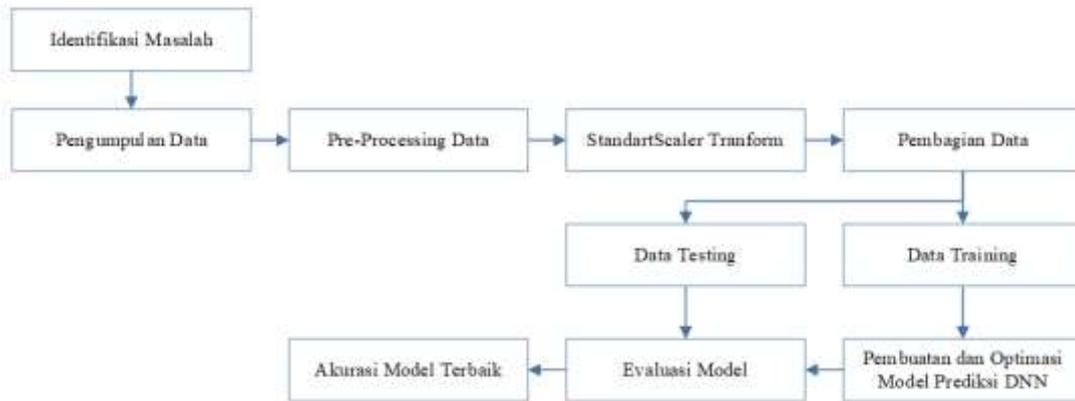
Perkembangan yang pesat dalam teknologi dan dinamika pasar global telah menciptakan lonjakan jumlah startup di berbagai sektor industri. Startup merupakan entitas bisnis yang terbentuk oleh sekelompok individu dengan tujuan mengembangkan produk atau layanan melalui pemanfaatan teknologi informasi [1]. Dalam kondisi ketidakpastian, startup memiliki tingkat risiko kegagalan yang signifikan karena produk atau layanan yang dikembangkan mungkin tidak sesuai dengan kebutuhan pengguna atau tidak mampu bersaing dengan perusahaan sejenis [2]. Keberhasilan suatu startup tidak hanya tergantung pada kebaruan ide yang diusungnya, melainkan juga pada kapasitas untuk mengelola dan menganalisis data guna membuat keputusan strategis yang efisien. Di era bisnis yang penuh dengan pertumbuhan startup yang pesat, peran dan dampaknya terhadap inovasi dan ekonomi semakin signifikan. Lonjakan jumlah startup mencerminkan kehidupan yang dinamis dalam ekosistem bisnis. Namun, seiring dengan pertumbuhan ini, hadir tantangan kompleks dan ketidakpastian yang dihadapi oleh startup. Secara keseluruhan, 90 dari 100 startup mengalami kegagalan dalam mencapai tingkat pendanaan yang diinginkan [3]. Keberhasilan suatu startup tidak hanya memiliki dampak pada kesuksesan bisnis individu tersebut, melainkan juga memberikan pengaruh luas terhadap dinamika ekonomi global. Oleh karena itu, dibutuhkan pengembangan model prediksi yang handal untuk memahami variabel-variabel yang mempengaruhi keberhasilan startup, sebagai dasar pengambilan keputusan yang baik dalam pengembangan startup.

Terdapat beberapa penelitian yang telah dilakukan sebelumnya dalam membangun model prediksi kesuksesan startup. Pada tahun 2021, Adhitya et al. melakukan penelitian prediksi kesuksesan startup dengan memanfaatkan *Decision Tree*, *K-NN*, dan *Naïve Bayes*, yang menghasilkan tingkat akurasi masing-masing sebesar 79,29%, 66,69%, dan 64,21% [4]. Dalam penelitian tersebut [4], jumlah variabel yang digunakan untuk pembuatan model prediksi sebanyak 19 dari total 49 variabel dalam *dataset*. Bangdiwala et al., membangun model prediksi startup menggunakan *Decision Trees*, *Random Forest*, *Gradient Boost*, *Logistic Regression*, dan *MLP Neural networks*. Hasil penelitian Bangdiwala et al., mendapatkan nilai akurasi sebesar 92% [5]. Ross et al., melakukan penelitian menggunakan algoritma *machine learning* untuk menghasilkan model prediksi startup dan mendapatkan nilai akurasi sekitar 80 – 89% [6]. Pada tahun 2022, Wulan & Halim menggunakan *Deep Neural Networks* untuk membangun model prediksi kesuksesan startup. Hasil penelitian menunjukkan bahwa model yang dihasilkan tidak mengalami *underfitting/overfitting* dengan tingkat akurasi sebesar 83,93% [7]. Untuk meningkatkan nilai akurasi dari penelitian [7], maka pada penelitian ini digunakan *StandartScaler Transform*.

StandartScaler Transform [8] melakukan transformasi pada *dataset* sehingga nilai rata-rata distribusinya menjadi nol dan deviasi standarnya menjadi satu. Nilai yang telah mengalami transformasi diperoleh dengan mengurangi nilai aslinya dari rata-rata dan membaginya dengan deviasi standar. Thara et, al. [9], menggunakan *StandartScaler Transform* untuk proses *feature scaling* pada model auto deteksi gejala kejang menggunakan *Deep Neural Networks*. Hasil percobaan yang dilakukan [9], diperoleh bahwa *StandartScaler Transform* menghasilkan akurasi yang lebih baik dibandingkan *RobustScaler*, *Normalizer*, dan *MinMaxScaler*.

Salah satu tahapan penting dalam mempersiapkan data sebelum diproses ke *Deep Neural Networks* adalah menerapkan *StandartScaler Transform*. Pendekatan ini berperan dalam normalisasi distribusi data, menciptakan dasar yang lebih optimal untuk *Deep Neural Networks* dalam memahami pola-pola yang kompleks. Oleh karena itu, penelitian ini berfokus pada integrasi *Deep Neural Networks* dan *StandartScaler Transform* dengan tujuan mengoptimalkan model prediksi kesuksesan *startup*. Melalui kombinasi *Deep Neural Networks* dan efektivitas *StandartScaler Transform*, diharapkan penelitian ini dapat memberikan sumbangan signifikan terhadap perkembangan metodologi prediksi dalam lingkungan *startup*, memberikan dampak positif pada pertumbuhan dan keberlanjutan ekosistem *startup*.

2. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Untuk menghasilkan model prediksi kesuksesan *startup* dengan nilai akurasi yang lebih baik, maka dilakukan tahapan penelitian seperti yang terlihat pada Gambar 1.

Berikut ini merukan penjelasan dari setiap tahapan yang dilakukan.

2.1 Identifikasi Masalah

Proses identifikasi masalah merupakan tahapan awal yang penting dilakukan untuk menganalisis apa yang menjadi masalah pada penelitian ini seperti yang telah dijabarkan pada bagian pendahuluan.

2.2 Pengumpulan Data

Pengumpulan data bertujuan untuk mengumpulkan data yang tepat sesuai dengan kebutuhan penelitian. Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari website <https://www.kaggle.com/datasets/manishkc06/startup-success-prediction>. *Dataset Startup Success Prediction* adalah kumpulan data yang mencakup perusahaan *startup* yang beroperasi di Amerika Serikat. Dalam *dataset* ini, terdapat 923 data yang mencakup 597 *startup* yang berhasil diakuisisi dan 326 *startup* yang telah ditutup. *Dataset* ini memiliki 49 variabel, dimana 48 variabel digunakan sebagai atribut, sedangkan 1 variabel (status) berfungsi sebagai kelas.

2.3 Pre-Processing Data

Berikut merupakan proses *pre-processing* yang dilakukan.

1. Pengecekan *Missing Value*

Missing value adalah kondisi di mana suatu variabel dalam data tidak memiliki nilai atau informasi yang lengkap, yang dapat disebabkan oleh kesalahan pengumpulan data atau ketidakhadiran responden. Penting untuk memeriksa dan menangani *missing value* sebelum melanjutkan ke proses analisis data agar mendapatkan data yang berkualitas.

2. Konversi Tipe Data

Pada tahap ini dilakukan proses konversi data-data kategorial *string* dan *object* menjadi data kategorial numerik sehingga dapat lebih mudah dipahami oleh model yang akan dibangun. Proses konversi dilakukan dengan menggunakan fungsi *label encoder*.

3. Pemilihan Variabel/Atribut

Pada tahap pemilihan variabel/atribut, pada penelitian ini digunakan *correlation matrix*. *Correlation matrix* merupakan metode statistik yang digunakan untuk mengevaluasi hubungan antara dua atau lebih variabel. Dalam *correlation matrix*, nilai korelasi berkisar antara -1 hingga 1 yang artinya jika dua buah variabel memiliki nilai 1, maka korelasi sempurna positif, yang berarti bahwa dua variabel bergerak bersamaan dalam arah yang sama. Jika nilai 0, maka tidak ada korelasi yang menunjukkan tidak ada

hubungan linier antara dua variabel. Sedangkan jika nilai -1 , maka korelasi sempurna negatif, yang berarti bahwa dua variabel bergerak berlawanan arah. Proses ini dilakukan untuk menghapus data atau variabel yang tidak relevan dengan model prediksi kesuksesan *startup*.

2.4 StandartScaler Transform

StandartScaler transform merupakan metode umum yang digunakan dalam *machine learning* untuk menghasilkan distribusi normal pada variabel-variabel numerik dengan nilai rata-rata 0 dan standart deviasi 1. Proses ini dilakukan untuk membantu mengoptimalkan model prediksi kesuksesan *startup* yang akan dibangun menggunakan *Deep Neural Networks* sehingga dapat menghasilkan hasil prediksi yang akurat.

2.5 Pembagian Data

Setelah proses *StandartScaler transform* selesai dilakukan, tahap selanjutnya adalah membagi dataset menjadi 70% data *training* dan 30% data *testing*. *Data training* akan digunakan saat proses pembuatan model prediksi kesuksesan *startup* menggunakan *Deep Neural Networks* dan data *testing* digunakan untuk evaluasi model yang telah dihasilkan.

2.6 Pembuatan dan Optimasi Model Prediksi Kesuksesan Startup

Proses selajutnya adalah pembuatan dan optimasi model prediksi kesuksesan *startup* menggunakan data *training* yang telah dipre-processing dan melewati proses *StandartScaler Transform*. Model *Deep Neural Networks* yang dihasilkan bertujuan untuk menghasilkan akurasi yang tinggi untuk memprediksi apakah sebuah *startup* akan sukses atau akan tutup.

2.7 Evaluasi Model

Evaluasi model dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* merupakan tabel evaluasi yang digunakan untuk melihat kinerja model klasifikasi pada *supervised learning* kedalam kategori yang benar atau salah. *Confusion matrix* yang digunakan dalam penelitian ini adalah akurasi dan presisi seperti terlihat pada tabel 1.

Tabel 1. *Confusion Matrix*

Actual Class	Assigned Class	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

$$\text{Akurasi} = (TP + TN) / (TP+TN+FP+FN) \dots\dots\dots(1)$$

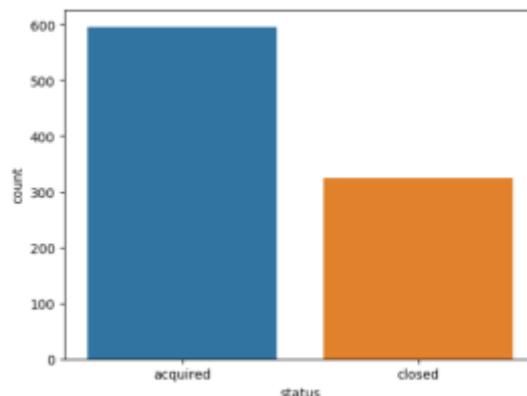
$$\text{Presisi} = (TP + TN) / (TP + FP) \dots\dots\dots(2)$$

3. HASIL DAN PEMBAHASAN

Berikut ini merupakan hasil dan pembahasan yang diperoleh dari penelitian ini.

3.1 Data Yang Digunakan

Pada penelitian ini menggunakan menggunakan *dataset* yang diperoleh dari website *Kaggle* [15] berisi 923 data yang mencakup 597 *startup* yang berhasil diakuisisi dan 326 *startup* yang telah ditutup seperti terlihat pada Gambar 2.



Gambar 2. Jumlah *Startup* Sukses dan *Closed*

3.2 Pre-Processing Data

Pada pre-processing data, proses pertama yang dilakukan adalah pengecekan *missing value*. Pada proses pengecekan *missing value*, ditemukan bahwa terdapat lima variabel atau atribut dengan *missing value*. Contohnya, variabel 'Unnamed:6' memiliki 53,41% *missing value*, 'closed_at' memiliki 63,71% *missing value*, 'age_first_milestone_year' memiliki 16,47% *missing value*, 'last_first_milestone_year' memiliki 16,47% *missing value*, dan 'state_code' memiliki 0,11% *missing value*. Setelah melakukan analisis, ditemukan bahwa variabel 'Unnamed:6', 'state_code', dan 'closed_at' tidak memiliki pengaruh signifikan terhadap kesuksesan *startup*, sehingga ketiga variabel tersebut dihapus. Selanjutnya, untuk 'age_first_milestone_year' dan 'last_first_milestone_year', nilai-nilai yang hilang diisi dengan 0 karena *startup* yang memiliki *missing value* pada kedua variabel tersebut belum memiliki *milestone*.

Setelah proses pengecekan *missing value*, selanjutnya dilakukan proses konversi tipe data untuk memastikan bahwa seluruh variabel objek/string yang digunakan dikonversi menjadi variabel numerik. Variabel 'category_code' merupakan variabel kategorial string/object yang akan diubah menjadi variabel kategorial numerik untuk memudahkan model memahami data dengan baik. Proses selanjutnya adalah dilakukan proses pemilihan variabel/atribut. Berdasarkan hasil analisis *Correlation matrix*, dipilih 36 variabel berikut yang akan digunakan dalam proses pembuatan model prediksi kesuksesan *startup*: 'latitude', 'longitude', 'labels', 'category_code', 'age_first_funding_year', 'age_last_funding_year', 'age_first_milestone_year', 'age_last_milestone_year', 'relationship', 'funding_rounds', 'funding_total_usd', 'milestones', 'is_CA', 'is_NY', 'is_MS', 'is_TX', 'is_otherstate', 'is_software', 'is_web', 'is_mobile', 'is_enterprise', 'is_advertising', 'is_gamesvideo', 'is_ecommerce', 'is_biotech', 'is_consulting', 'is_othercategory', 'has_vc', 'has_angel', 'has_roundA', 'has_roundB', 'has_roundC', 'has_roundD', 'avg_participants', 'is_top500', dan 'status'.

3.3 StandartScaler Transform

StandartScaler merupakan teknik pre-processing data yang berfungsi untuk mengubah variabel/atribut numerik pada *dataset* sehingga memiliki nilai *mean* = 0 dan standart deviasi = 1. Proses yang dilakukan pada tahap ini adalah StandartScaler menghitung *mean* dan standart deviasi dari setiap variabel/atribut dalam *dataset*. Setelah itu dilakukan transformasi data untuk menghasilkan nilai *mean* =0 dan standart deviasi = 1. Proses tranformasi data ini penting untuk membantu menyeragamkan seluruh variabel/atribut dalam *dataset* yang dapat membantu algortima *Deep Neural Networks* untuk dapat menghasilkan model prediksi yang lebih akurat.

3.4 Pembuatan dan Optimasi Model Prediksi

Untuk membangun dan optimasi model prediksi digunakan dua skenario berikut ini:

1. Pembuatan model prediksi kesuksesan *startup* menggunakan *Deep Neural Networks*
2. Pembuatan model prediksi kesuksesan *startup* menggunakan *Deep Neural Networks* dengan StandartScaler Tranform

Pembuatan model dilakukan dengan metode ekperimen untuk mendapatkan akurasi terbaik. Berdasarkan hasil ekperimen yang telah dilakukan, untuk membangun model *Deep Neural Networks* digunakan parameter berikut:

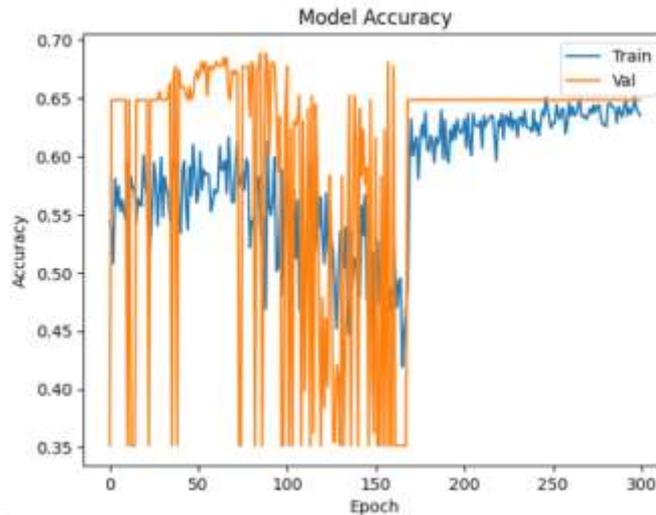
1. *Input layer* dengan jumlah sebanyak 35 sesuai dengan jumlah variabel/atribut yang digunakan.
2. 4 *hidden layer* dengan masing-masing 100 *neuron*
3. 1 *output layer* untuk menghasilkan luaran apakah *startup* akan *acquired* atau *closed*
4. *Relu* digunakan sebagai fungsi aktivasi pada *input* dan *hidden layer* sedangkan *sigmoid* digunakan sebagai fungsi aktivasi pada *output layer*
5. Adam digunakan sebagai fungsi optimasi model
6. Jumlah *epoch* = 300

Pembuatan model dilakukan dengan menggunakan 70% data *training* yang telah diproses sebelumnya. Tabel 2 merupakan hasil pembuatan model dengan 2 skenario yang telah disebutkan di atas.

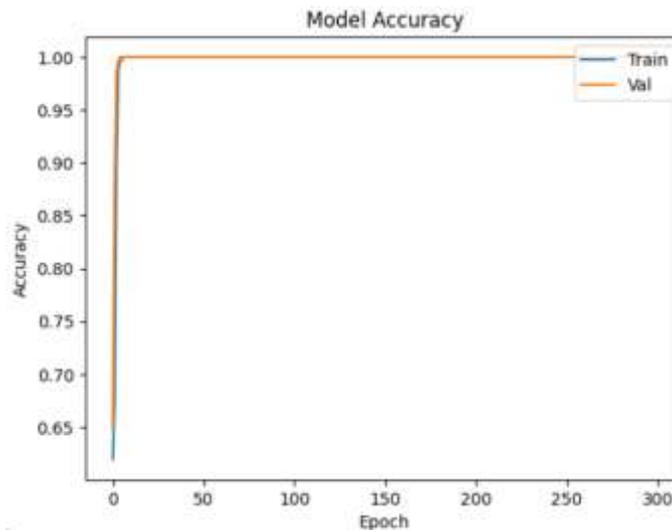
Tabel 2. Hasil Pembuatan Model Prediksi

Metode	Variabel/Atribut	Akurasi (%)	Presisi (%)
DNN	35	64.75	0
DNN + StandartScaler Transform	35	100	100

Gambar 3, merupakan hasil akurasi pembuatan model prediksi dengan *Deep Nerual Networks* dan Gambar 4, merupakan hasil akurasi pembuatan model prediksi dengan *Deep Neural Networks* dengan StandartScaler Transform.



Gambar 3. Hasil Akurasi Model Prediksi Dengan *Deep Neural Networks*



Gambar 4. Hasil Akurasi Model Prediksi Dengan *Deep Neural Networks* dan *StandartScaler Transform*

Berdasarkan Gambar 3 dapat dilihat bahwa model prediksi dengan *Deep Neural Networks* mengalami *underfitting/overfitting* sedangkan Gambar 4, menunjukkan bahwa model prediksi dengan *Deep Neural Networks* dan *StandartScaler Transform* sudah *bestfitting* yang artinya model dapat dilanjutkan ke proses evaluasi dengan data *testing*. Pada tahap *training*, terjadi peningkatan akurasi sebesar 35,25%.

3.5 Evaluasi dan Pembahasan

Model yang telah dibangun selanjutnya di evaluasi menggunakan 30% data *testing*. Tabel 3 merupakan hasil evaluasi model.

Tabel 3. Hasil Evaluasi Model Prediksi

Metode	Variabel/Atribut	Akurasi (%)	Presisi (%)
DNN	35	63.04	0
DNN + StandartScaler Transform	35	100	100

Berdasarkan hasil pengujian yang telah dilakukan pada Tabel 3, diperoleh bahwa akurasi model prediksi kesuksesan *startup* dengan *Deep Neural Networks* dan *StandartScaler Tranform* lebih baik dibandingkan dengan hanya menggunakan *Deep Neural Networks*.

4. SIMPULAN

Eksperimen dan evaluasi yang telah dilaksanakan menunjukkan bahwa optimalisasi model prediksi kesuksesan *startup* melalui *Deep Neural Networks* berhasil ditingkatkan dengan pemanfaatan *StandartScaler Transform*. Selama proses pengembangan model, terjadi peningkatan signifikan pada tingkat akurasi sekitar 35,25%, dan pada tahap pengujian, tercatat peningkatan akurasi sebesar 36,96%. Model prediksi kesuksesan

startup yang diajukan berhasil menghindari masalah *overfitting/underfitting*. Oleh karena itu, model prediksi ini dapat diandalkan dan berpotensi menjadi dasar yang kuat untuk pengembangan aplikasi prediksi kesuksesan *startup*.

5. SARAN

Adapun saran-saran yang dapat diberikan untuk penelitian selanjutnya adalah:

1. Untuk penelitian selanjutnya model ini dapat diterapkan untuk dataset startup pada kategori tertentu saja untuk menguji apakah model prediksi yang dibangun dapat menghasilkan nilai akurasi yang tinggi atau tidak.
2. Model prediksi yang dibangun juga dapat diuji dengan menggunakan dataset yang lain seperti dataset dari *angelist.co* dan lain-lain untuk memprediksi performance sebuah startup.

UCAPAN TERIMA KASIH

Peneliti mengungkapkan apresiasi kepada Universitas Mikroskil atas dukungan yang diberikan selama pelaksanaan penelitian ini.

DAFTAR PUSTAKA

- [1] Y. Birkman, “Hello, Startup: A Programmer's Guide to Building Products, Technologies, and Teams,” O'Reilly Media, Inc., 2015.
- [2] Goldenia, et al., “Implementasi Algoritma Support Vector Machine dalam Memprediksi Keberhasilan Suatu Startup Berdasarkan Status Akuisisi,” Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA), 2021.
- [3] E. Ramalakshmi, and S. R. Kamidi, “Predictions for Startups”, International Journal of Engineering & Technology, vol. 7, 2018.
- [4] A. P. Adhitya, K. Ainiyah, and K. F. H. Holle, “Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up”, JISKA, vol. 6, No. 3, Pp. 178 – 188, 2021.
- [5] M. Bangdiwala, Y. Mehta, S. Agrawal and S. Ghane, “Predicting Success Rate of Startups using Machine Learning Algorithms,” 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, 2022, pp. 1-6, doi: 10.1109/ASIANCON55314.2022.9908921.
- [6] G. Ross, S. Das, D. Sciro, and H. Raza, “CapitalVX: A machine learning model for startup selection and exit prediction,” The Journal of Finance and Data Science, vol. 7, Pp. 94-114, 2021.
- [7] W. S. Lestari & A. Halim, “Prediksi Kesuksesan Startup Menggunakan Deep Neural Network,” Jurnal SIFO Mikroskil, vol. 23, no. 2, 2022.
- [8] scikit-learn developers. (2023, Des) Compare the effect of different scalers on data with outliers [online]. Available: https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html#standardscaler
- [9] T. D. K. Thara, B. G. P. Sudha and F. Xiong, “Auto-detection of Epileptic Seizure Events Using Deep Neural Network with Different Feature Scaling Techniques,” Pattern Recognition Letters, 2019.