

Implementasi Algoritma *Support Vector Machine* dalam Memprediksi Keberhasilan Suatu *Startup* Berdasarkan Status Akuisisi

Audrey Era Goldenia¹, Cindy Chairunnisa², Hilda Harisa³, Johannes Christian⁴ dan Desta Sandya Prasvita⁵
Informatika / Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jakarta
Jl. RS. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia
audreyeg@upnvj.ac.id¹, cindyc@upnvj.ac.id², hildah@upnvj.ac.id³, johannesc@upnvj.ac.id⁴,
desta.sandya@upnvj.ac.id⁵

Abstrak. Startup merupakan sebuah perusahaan yang dibangun untuk menjual produk ke seluruh market dengan cara yang baru. Pertumbuhan Startup dalam beberapa tahun terakhir mengalami peningkatan di dunia. Namun, dari banyaknya perusahaan Startup yang ada tidak semua mengalami keberhasilan. Terdapat beberapa faktor yang mempengaruhi keberhasilan dari perusahaan startup diantaranya seperti efisiensi perencanaan, efisiensi penggunaan dana, kerjasama tim, biaya, dan sebagainya. Penelitian ini akan menggunakan metode Support Vector Machine (SVM) untuk memprediksi keberhasilan suatu startup berdasarkan status akuisisi perusahaan startup tersebut. Dari hasil uji coba yang telah dilakukan, terbukti bahwa metode Support Vector Machine (SVM) dapat digunakan untuk menentukan keberhasilan perusahaan startup dengan akurasi sebesar 79.1% dengan nilai hyperplane Kernel : Linear & C : 1.0.

Kata Kunci: Prediksi, *Startup*, SVM.

1 Pendahuluan

Semakin berkembangnya kemajuan teknologi yang memberikan dampak ke berbagai bidang kehidupan. Salah satu pemanfaatan perkembangan teknologi yang dilakukan pada saat ini adalah dengan membangun sebuah perusahaan rintisan atau Startup. Pembangunan sebuah perusahaan Startup dalam perekonomian dinilai mampu meningkatkan antusiasme kewirausahaan para wirausahawan untuk terus tumbuh dan saling bersaing dalam bertahan hidup [1]. Sehingga pertumbuhan Startup di dunia selama beberapa tahun terakhir mengalami kenaikan. Berdasarkan data yang terdapat pada StartupRanking (4/6/2021), negara dengan jumlah perusahaan startup terbesar adalah negara Amerika Serikat dengan jumlah perusahaan sebesar 99.126 startup. Banyaknya perusahaan startup yang ada tidak semuanya mengalami keberhasilan. Adapun faktor - faktor yang mempengaruhi keberhasilan dari perusahaan startup diantaranya seperti efisiensi perencanaan, efisiensi penggunaan dana, kerjasama tim, biaya, dan sebagainya[2].

Berdasarkan penelitian pada jurnal yang telah dilakukan sebelumnya untuk melakukan prediksi keberhasilan perusahaan startup dengan menggunakan salah satu metode machine learning yaitu Random Forest. Metode yang digunakan tersebut menghasilkan nilai akurasi maksimum mencapai 87% [2]. Selain itu, pada penelitian lainnya, metode Support Vector Machine (SVM) digunakan untuk memprediksi retensi mahasiswa dengan hasil tingkat akurasi mencapai 90,4% [3]. Penelitian lain juga pernah dilakukan untuk memprediksi keberhasilan startup dengan metode Support Vector Machine dan Random Forest dan mendapatkan nilai AUC berturut-turut yaitu 0.51 dan 0.61[11]. Algoritma SVM juga kerap kali digunakan untuk memprediksi keberhasilan model bisnis dengan akurasi 83.6%[12].

Pada penelitian ini akan dilakukan prediksi keberhasilan suatu perusahaan startup dengan menggunakan metode klasifikasi Support Vector Machine (SVM). Metode SVM bekerja dengan melakukan klasifikasi suatu variabel input ke dalam kelas yang diharapkan dengan membuat hyperplane diantaranya kemudian memaksimalkan perbedaan nilai antara kelas dan hyperplane yang berbeda pada setiap kesalahan dalam klasifikasi sehingga dapat meminimalkan kesalahan dalam melakukan klasifikasi [4]. Penelitian ini juga akan melakukan perbandingan hasil

akurasi terhadap model Random Forest dan Support Vector Machine (SVM) dalam melakukan prediksi keberhasilan suatu perusahaan startup.

2 Landasan Teori

2.1 Startup

Awal mula munculnya istilah “start up” berkaitan dengan beberapa perusahaan mikro yang muncul, khususnya perusahaan yang aktif dibidang teknologi dan komputer [6]. Secara umum, startup merupakan bisnis yang dikembangkan dan dibangun oleh seorang pengusaha dengan memanfaatkan ide baru dan sumber daya bisnis [7]. Definisi lain, startup adalah sebuah perusahaan yang dibangun oleh seorang wirausahawan yang menjual produk ke seluruh area market dengan cara yang belum pernah dilakukan oleh perusahaan-perusahaan sebelumnya. Atau dapat dikatakan startup merupakan perusahaan yang dibuat untuk mengembangkan suatu produk atau layanan baru dalam kondisi yang tidak pasti. Sehingga pembangunan perusahaan startup memiliki resiko yang tinggi dan tidak dapat diprediksi keberhasilannya karena produk yang dijual mungkin tidak bekerja dengan baik oleh penggunanya dan membutuhkan penyesuaian. Perusahaan startup umumnya berkaitan dengan teknologi yang dijadikan sebagai sebuah produk atau layanan [5].

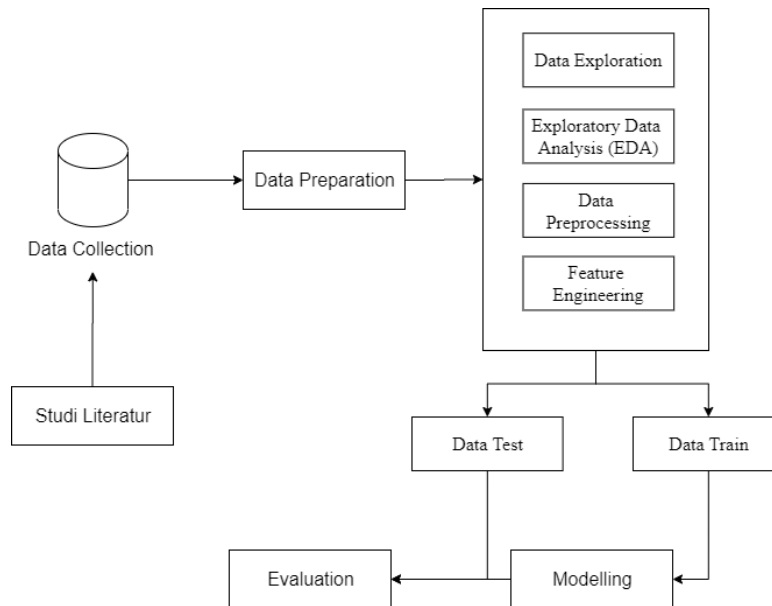
Seorang pencetus “Lean Startup Movement” atau dikenal dengan Steve Blank mendeskripsikan bahwa startup merupakan suatu organisasi yang bersifat sementara dan memiliki tujuan untuk mengembangkan model bisnis yang terukur serta melakukannya dengan waktu yang singkat dan memiliki dampak untuk pasar[8,9]. Untuk pemahaman yang lebih baik terkait definisi startup terdapat tiga poin utama yang perlu diketahui, diantaranya perusahaan startup memiliki tujuan untuk mengisi kekosongan dari kebutuhan pasar yang belum terpenuhi, startup melakukan hipotesis model bisnis secara berulang hingga hipotesis tersebut terbukti, dan pendiri dari startup harus melakukan validasi terkait model dari hipotesis dan memeriksa sikap dari pelanggan apakah sesuai dengan apa yang diprediksikan[9].

Dalam prosesnya, startup memiliki siklus atau tahapan kegiatan yang menunjukkan perkembangan dari startup itu sendiri. Tahapannya antara lain terdiri dari ideation, intention, startup, dan expansion[10]. Tahapan ideation merupakan tahap yang berfokus untuk menemukan dan membuat ide yang memiliki potensi untuk berkembang. Ide yang dicari diharapkan dapat memecahkan suatu masalah yang ada dan berguna untuk pelanggan. Tahapan selanjutnya yaitu intention dimana pada tahap ini pengusaha mengeksekusi idenya ke dalam bentuk bisnis dengan melakukan berbagai kegiatan yang mendukung. Tahap intention terlihat dari adanya motivasi, kepercayaan diri, kecenderungan, dan resiko yang harus dihadapi. Selanjutnya, tahapan startup yaitu pengusaha telah meluncurkan bisnisnya ke pasar. Terakhir, tahap expansion yaitu pengusaha telah berhasil mengembangkan bisnisnya menjadi lebih besar. Pada tahap ini perlu dikembangkan keterampilan dan keahlian baru agar bisnis startup yang dijalankan memiliki inovasi yang berkembang[10].

2.2 Support Vector Machine (SVM)

SVM adalah algoritma pembelajaran terawasi yang dapat melakukan klasifikasi atau regresi untuk kategoris dan variabel respon numerik, masing-masing. Algoritma ini menciptakan ruang pemetaan untuk memisahkan data input di kelas yang berbeda. Model ini mampu memetakan data linier dan non-linier dengan menerapkan fungsi kernel yang dapat mengubah input ke ruang dimensi yang lebih tinggi, yang memungkinkan pemisahan linier. Kemudian, penggunaan kernel mengurangi kompleksitas masalah dengan membuat hyperplane paralel yang memisahkan data. Kondisi optimal ditemukan dengan meminimalkan norma Euclidean dari vektor bobot, yang merupakan masalah optimasi terkendala yang dapat diselesaikan dengan menggunakan metode pengali LaGrange. Algoritma memaksimalkan margin antara paralel margin hyperplanes membatasi kesalahan klasifikasi. Diasumsikan bahwa dengan bertambahnya jarak antara hyperplanes kesalahan generalisasi berkurang. Salah satu keuntungan menggunakan SVM adalah dapat bekerja dengan baik dengan data sampel kecil [6].

3 Metode Penelitian



Gambar. 1. Bagan alur.

3.1 Studi Literatur

Sebelum melakukan penelitian dilakukan studi literatur untuk mencari berbagai informasi yang berhubungan dengan penelitian. Sumber informasi didapatkan dari berbagai sumber seperti jurnal, buku, artikel, website dan sebagainya. Studi literatur ini dilakukan untuk menentukan topik penelitian yang akan dikerjakan. Berdasarkan studi literatur yang telah dilakukan, topik yang akan digunakan dalam penelitian ini adalah terkait suatu perusahaan startup.

Berdasarkan data pertumbuhan startup dari tahun ke tahun mengalami pertumbuhan dan semakin meningkat jumlahnya di berbagai negara. Oleh sebab itu, setelah dilakukan analisa terkait pertumbuhan startup ini akan dilakukan prediksi untuk menentukan kesuksesan suatu perusahaan startup.

3.2 Data Collection

Pada penelitian ini data yang akan digunakan merupakan sebuah data sekunder yang diperoleh dari kaggle.com. Dataset yang berjudul Startup Success Prediction merupakan dataset dari kumpulan perusahaan startup yang berada di negara Amerika. Dataset ini memiliki jumlah 923 record data yang terdiri dari 49 variabel, dimana diantaranya 48 variabel sebagai atribut dan 1 variabel lainnya sebagai label atau kelas.

3.3 Data Preparation

Tahap ini merupakan salah satu tahap yang sangat penting dalam sebuah penelitian sebelum dilakukan pembuatan model. Data yang telah didapatkan sebelumnya masih memiliki beberapa data yang hilang, tidak sesuai, terduplikasi dsb. dan perlu dilakukan beberapa tahapan preprocessing dari data tersebut sehingga data yang akan digunakan untuk membangun model merupakan data yang benar-benar berkualitas dan menghasilkan nilai akurasi model yang baik.

4 Hasil dan Pembahasan

Data yang digunakan dalam penelitian merupakan data yang berjumlah 923 data startup di Amerika. Tabel 1 berikut menunjukkan rincian data startup yang masih mentah atau belum dilakukan pengolahan.

Tabel 1. Dataset Startup yang belum dikelola.

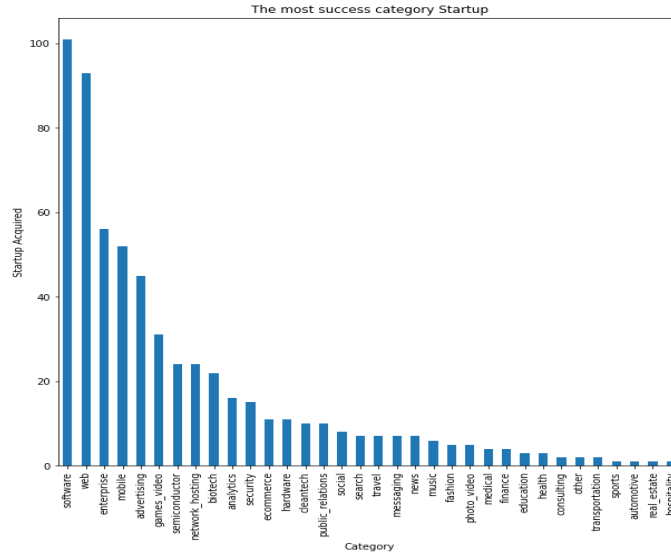
No	Unna med: 0	state_ code	latitu de	longit ude	zip_c ode	id	city	Unname d: 6	name	...	status
1	1005	CA	42.35 888	- 71.05 7	9210 1	c:66 69	San Diego		Bandsi ntown	...	acquired
2	204	CA	37.23 8916	- 121.9 7	9503 2	c:16 283	Los Gatos		TriCip her	...	acquired
3	1001	CA	32.90 1049	- 117.1 9	9212 1	c:65 620	San Diego	San Diego CA 92121	Plixi	...	acquired
4	738	CA	37.32 0309	- 122.0 5	9501 4	c:42 668	Cuper tino	Cupertin o CA 95014	Solidco re System s	...	acquired
5	1002	CA	37.77 9281	- 122.4 2	9410 5	c:65 806	San Franc isco	San Francisc o CA 94105	Inhale Digital	...	closed
...
923	462	CA	37.38 6778	- 121.9 7	9505 4	c:26 702	Santa Clara	Santa Clara CA 95054	Asemp ra Techno logies	...	acquired

4.1 Data Exploration

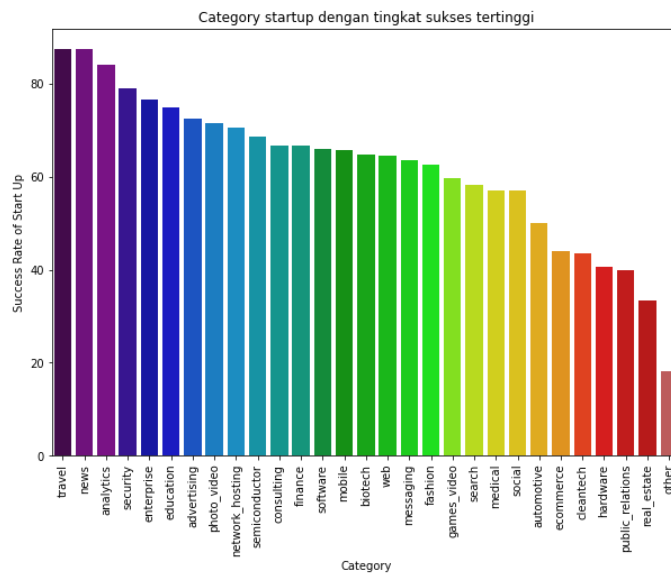
Berdasarkan dataset startup yang digunakan, agar lebih mudah dalam pengelolaan dilakukan perubahan bentuk nilai pada variabel "status" dimana acquired diubah menjadi 1 dan closed menjadi 0. Selain itu, ketika dilakukan pengecekan ditemukan adanya kesamaan antara kolom variabel "label" dengan "status" sehingga perlu dilakukan penghapusan pada salah satu kolom yakni kolom "label". Pada kolom lainnya, antara kolom "state_code" dan "state_code.1" hanya memiliki satu value yang berbeda, sehingga dilakukan penghapusan pada kolom "state_code.1".

4.2 Exploratory Data Analysis (EDA)

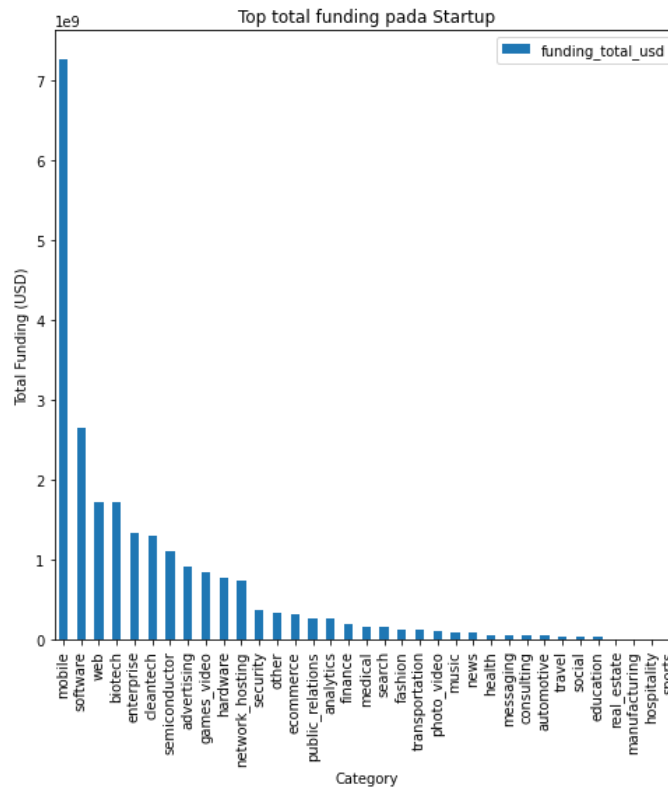
Setelah dilakukan analisa data, berdasarkan data startup yang digunakan dapat diambil beberapa gambaran baru untuk perkembangan startup di Indonesia kedepannya sebagaimana yang tertera pada Gambar 2, 3, dan 4 di bawah ini.



Gambar. 2. Kategori startup dengan jumlah keberhasilan tertinggi.



Gambar. 3. Kategori startup dengan tingkat kesuksesan tertinggi.



Gambar. 4. Kategori startup dengan funding terbesar.

Berdasarkan Gambar 2. ditunjukkan bahwa kategori startup yang memiliki jumlah keberhasilan tertinggi yaitu kategori software dilanjutkan web dan enterprise. Untuk kategori dengan jumlah keberhasilan terendah yaitu pada kategori hospitality. Selanjutnya, pada Gambar 3. ditunjukkan bahwa kategori startup dengan tingkat keberhasilan tertinggi yaitu pada kategori travel dan terendah pada kategori lainnya. Terakhir, pada Gambar 4. ditunjukkan bahwa funding terbesar dimiliki startup dengan kategori mobile diikuti software dan kemudian web. Sedangkan, funding terendah berada pada kategori sports.

4.3 Data Preprocessing

Setelah dilakukan eksplorasi dan analisis data, berdasarkan data startup yang kami dapatkan, setelah dilakukan pengidentifikasian missing value didapatkan hasil pada Tabel 2.

Tabel 2. Identifikasi missing values

Features	Null Values	Missing (%)
Unnamed:6	493	53.41
closed_at	588	63.7
age_first_milestone_year	152	16.46
age_last_milestone_year	152	16.46

Berdasarkan Tabel 2 terlihat bahwa pada fitur “Unamed:6” dan “closed_at” memiliki persentase missing value diatas 50%. Fitur “Unamed:6” merupakan fitur gabungan dari beberapa fitur diantaranya city, state_code, dan zip_code. Untuk itu, kami melakukan penghapusan kolom “Unamed:6” dan kemudian diisi kembali dengan gabungan dari fitur city, state_code, dan zip_code. Pada fitur “closed_at” yang merupakan data tanggal penutupan suatu startup, hal ini tidak mendukung dalam pemodelan sehingga dilakukan penghapusan kolom. Selanjutnya, missing value pada variabel age_first_milestone_year dan age_last_milestone_year disebabkan startup tersebut belum ada milestones sehingga kami mengisi missing value dengan 0.

Selanjutnya dilakukan pengecekan terkait adanya duplikasi dalam data. Dalam pengecekan tidak ditemukan adanya duplikasi data, namun sebelumnya pada eksplorasi data telah dilakukan analisis kategorikal dan ditemukan fitur yang memiliki kesamaan record data diantaranya name, id, dan object_id. Sehingga, untuk mengatasinya kami melakukan penghapusan pada ketiga fitur tersebut. Duplikasi data ditujukan pada Tabel 3.

Tabel 3. Identifikasi duplikasi data.

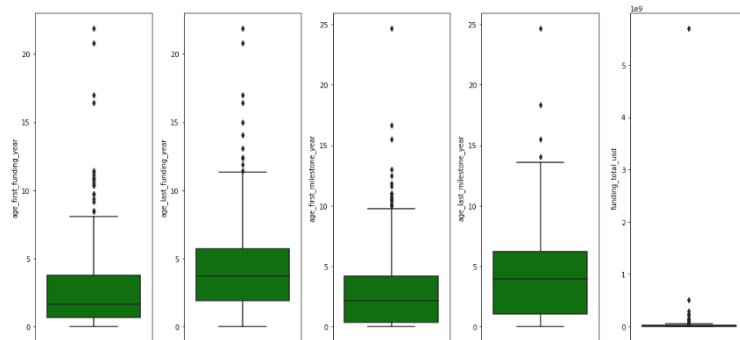
No	Unna med: 0	state_co de	latitud e	longi tude	zip_c ode	id	city	Unnamed: 6	name	...	status
832	505	CA	37.481 51	- 121.9 4532 8	94538	c:28 482	Freemo nt	Fremont CA 94538	Redwood Systems	...	1

Selain itu, dilakukan pengecekan terkait value yang bernilai negatif pada variabel yang seharusnya tidak memiliki nilai negatif. Dalam pengecekan ditemukan adanya value yang bernilai negatif pada kolom age_first_funding_year, age_last_funding_year, age_first_milestones_year, dan age_last_milestones_year. Untuk mengatasi hal tersebut, kami melakukan penghapusan pada kolom yang memiliki nilai negatif. Negative values pada data ditujukan pada Tabel 4.

Tabel 4. Identifikasi negative values

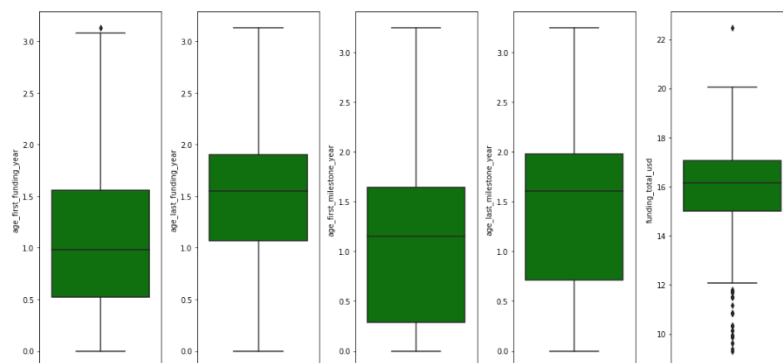
Features	Negative Values
age_first_funding_year	46
age_last_funding_year	13
age_first_milestone_year	46
age_last_milestone_year	12

Selanjutnya dilakukan pengecekan outlier pada data kemudian ditemukan adanya outlier pada fitur-fitur numerik diantaranya age_first_funding_year, age_last_funding_year, age_first_milestone_year, age_last_milestone_year, dan funding_total_usd. yang ditunjukkan pada Gambar 5.



Gambar. 5. Outlier pada fitur numerik.

Sehingga untuk mengatasi outlier pada fitur-fitur tersebut, kami menggunakan log-transformation pada tiap fitur sehingga terlihat perubahan pada Gambar 6.



Gambar. 6. Outlier pada fitur numerik setelah dilakukan log-transformation.

4.4 Feature Engineering

Selanjutnya dilakukan penambahan fitur-fitur yang akan digunakan untuk mendukung pemodelan. Pada data tersebut terdapat fitur mengenai funding atau pendanaan diantaranya Round A, Round B, Round C dan Round D. Kemudian dilakukan penambahan fitur yaitu `has_RoundABCD` untuk melihat apakah perusahaan startup tersebut memiliki funding A, B, C atau D. Dan untuk menentukan apakah suatu perusahaan startup memiliki seorang investor atau tidak dilakukan penambahan fitur yaitu `has_Investor` dengan menggunakan hubungan antara fitur `has_VC` dan `has_angel` yang memiliki nilai data bernilai 1.

Setelah dilakukan penambahan kedua fitur dilakukan penambahan fitur baru dengan menggunakan hubungan antara kedua fitur tersebut yaitu fitur `has_Seed`. Fitur ini merupakan fitur yang menentukan ada atau tidak adanya pendanaan pada sebuah perusahaan startup yang akan mempengaruhi pertumbuhan dan kesuksesan suatu perusahaan startup. Sehingga proses penambahan fitur pada penelitian ini merupakan salah satu proses penting untuk memprediksi keberhasilan suatu perusahaan startup.

4.5 Modelling

Setelah data yang akan digunakan telah dilakukan preprocessing maka selanjutnya adalah melakukan pembagian data. Data dibagi menjadi data training dan data testing dengan perbandingan 80% sebagai data training dan 20% sebagai data testing. Pembagian data ini menghasilkan pembagian untuk data training sejumlah 671 data dan data testing sejumlah 168 data.

Selanjutnya setelah memperoleh data yang akan digunakan untuk melakukan pengujian model klasifikasi Support Vector Machine (SVM) kemudian dilakukan percobaan perbandingan nilai hyperparameters dan hyperplane untuk mendapatkan model klasifikasi dengan nilai akurasi yang terbaik. Perbandingan nilai hyperparameters dan hyperplane yang digunakan pada data training yaitu Kernel : Linear & C : 1.0, Kernel : Linear & C : 100.0, Kernel : rbf & C : 1.0, Kernel : rbf & C : 100.0, Kernel : Poly & C : 1.0 dan Kernel : Poly & C : 100.0.

4.3 Evaluasi

Selanjutnya untuk menentukan keakuratan model klasifikasi yang telah dibentuk dengan beberapa perbandingan nilai hyperplane adalah menentukan akurasi dari masing-masing model yang telah dibentuk. Tingkat akurasi menggambarkan seberapa baik model dalam mengklasifikasi data. Dalam penelitian ini, nilai akurasi dilakukan dengan menggunakan evaluasi model Confusion Matrix. Perhitungan tabel confusion matrix direpresentasikan dengan 4 nilai klasifikasi, yaitu : True Positive (TP), True Negative (TN), False Positif (FP) dan False Negative (FN). Hasil dari akurasi, presisi dan error pada percobaan model klasifikasi yang telah dibentuk ditampilkan pada Tabel 5.

Tabel 5. Hasil akurasi berdasarkan nilai hyperplane

Kernel	C	Akurasi	Presisi	Error
Linear	1.0	79.17%	79.41%	20.83%
Linear	100.0	77.38%	79.39%	22.62%
Rbf	1.0	75.6%	76.06%	24.4%
Rbf	100.0	71.43%	78.63%	28.57%
Polynomial	1.0	71.43%	74.81%	28.57%
Polynomial	100.0	70.24%	79.28%	29.76%

Berdasarkan Tabel 5 di atas telah didapatkan model klasifikasi dengan nilai akurasi terbaik yaitu dengan nilai hyperplane Kernel : Linear & C : 1.0. Model klasifikasi yang dibentuk untuk melakukan klasifikasi suatu perusahaan startup pada penelitian ini telah mendapatkan nilai akurasi sebesar 79.17%.

5 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan pada data startup maka didapatkan kesimpulan terhadap prediksi keberhasilan suatu startup adalah sebagai berikut :

1. Dihilangkan 34 variabel pada data perusahaan startup yang memiliki pengaruh yang besar untuk menentukan keberhasilan perusahaan startup.
2. Dilakukan perbandingan beberapa nilai hyperplane pada model klasifikasi sehingga model klasifikasi yang didapatkan mendapatkan nilai akurasi yang optimal.
3. Hasil klasifikasi terhadap data perusahaan startup dengan model klasifikasi Support Vector Machine (SVM) memperoleh nilai akurasi sebesar 79.1% dengan nilai hyperplane Kernel : Linear & C : 1.0.

Referensi

- [1] Boyoung, Kim; Hyojin, Kim; Youngok, Jeon. Critical Success Factors of a Design Startup Business. 2018. Sustainability.
- [2] Ramalakshmi, Eliganti; Kamidi, Sindhuja Reddy. Predictions for Startups. 2018. International Journal of Engineering & Technology.
- [3] Cardona, Tatiana A; Cudney, Elizabeth A. Prediction Student Retention Using Support Vector Machine. 2019. Chicago : Elsevier Ltd.
- [4] Haykin, Simon. Neural Networks and Learning Machines. 2009. New York: Prentice Hall.
- [5] Ries, E. The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. 2011. Crown Publishing Group: New York, NY, USA.
- [6] Skawinska, Eulalia; Zalewski, Romuald I. Success Factors of Startups in the EU – A Comparative Study. 2020. Poland. Sustainability.
- [7] Aruan, Carina; Worang, Frederik G; Saerang, Regina T. Analyzing The Use Of Hyperlocal Marketing Of Street Food Startup Business In Manado. 2019. Jurnal EMBA.
- [8] Blank, Steve; Euchner Jim. The Genesis and Future of Lean Startup: An Interview with Steve Blank. 2018. Research Technology Management.
- [9] Roshanzamir, Amirhossein; Roosta, Ahmad; Shajari, Mehdi. Startups: Challenges and Implications to Success. 2019. World of Export.
- [10] Quinto, Ivana; Rippas, Pierluigi; Thomas, Anthonio; Passaro, Renato. The Start-up Lifecycle: An Interpretative Framework Proposal. 2016. Conference: XVII Annual Scientific Meeting of the Italian Association of Management Engineering (AiIG).
- [11] Jinze Li. Prediction of the Success of Startup Companies Based on Support Vector Machine and Random Forest. 2020. WAIE: Canada.
- [12] Bohm, Markus dkk. The Business Model DNA: Towards an Approach for Predicting Business Model Success. 2017. Wirtschafts Informatik: Germany.