ISSN: 2598-0256, E-ISSN: 2598-0238

Kajian Perbandingan Algoritma KNN Dan SVM Untuk Prediksi Pengangguran Di Provinsi Lampung

Ganes Angga Witata[1a], Joko Triloka[2b]

[a b]Institut Informatika & Bisnis Darmajaya
[a]*ganes.2221210012@mail.darmajaya.ac.id*
[a]*joko.triloka@darmajaya.ac.id*

Abstract

Unemployment is a complex social and economic problem in Indonesia, including in Lampung Province, which has a negative impact on poverty and economic imbalance. Indonesia has the second highest unemployment rate in Southeast Asia. Lampung Province also faces unemployment challenges with Bandar Lampung City having the highest in the province. Identifying the factors that influence unemployment and developing an effective classification method is essential. This study aims to compare the performance of the K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) algorithms in classifying unemployment in Lampung Province. This research contributes to a better understanding of the factors that influence the unemployment rate and provides recommendations about the most effective algorithms. This study uses modeling and data analysis techniques that are commonly used in various disciplines to overcome these problems. The KNN algorithm classifies based on similarity to nearest neighbors, while SVM uses a kernel-based or linear separation approach. The findings of this study will enhance our understanding of the factors of unemployment and support decision making in addressing the problem.

Keywords: Unemployment, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Classification, Lampung Province.

Abstrak

Pengangguran adalah masalah sosial dan ekonomi yang kompleks di Indonesia, termasuk di Provinsi Lampung.yang berdampak negatif kemiskinan dan ketidakseimbangan ekonomi. Indonesia memiliki tingkat pengangguran kedua tertinggi di Asia Tenggara. Provinsi Lampung juga menghadapi tantangan pengangguran dengan Kota Bandar lampung memiliki tertinggi di provinsi tersebut. Mengidentifikasi faktor-faktor yang memengaruhi pengangguran dan mengembangkan metode klasifikasi yang efektif sangat penting. Penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Nearest Neighbors (KNN) dan Support Vector Machines (SVM) dalam mengklasifikasikan pengangguran di Provinsi Lampung. Penelitian ini memberikan kontribusi untuk pemahaman yang lebih baik tentang faktor-faktor yang memengaruhi tingkat pengangguran dan memberikan rekomendasi tentang algoritma yang paling efektif. Penelitian ini menggunakan teknik pemodelan dan analisis data yang umum digunakan dalam berbagai disiplin ilmu untuk mengatasi masalah tersebut. Algoritma KNN mengklasifikasikan berdasarkan kesamaan dengan tetangga terdekat, sedangkan SVM menggunakan pendekatan pemisahan linear atau berbasis kernel. Temuan penelitian ini akan meningkatkan pemahaman kita tentang faktor-faktor pengangguran dan mendukung pengambilan keputusan dalam mengatasi masalah tersebut.

Kata Kunci: Pengangguran, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Klasifikasi, Provinsi Lampung.

1. PENDAHULUAN

Pengangguran adalah masalah kompleks di Indonesia, termasuk di Provinsi Lampung, dengan dampak negatif seperti kemiskinan dan ketidakseimbangan ekonomi. Pada tahun 2023, Indonesia memiliki tingkat pengangguran kedua tertinggi di Asia Tenggara. Provinsi Lampung juga menghadapi tantangan pengangguran, dengan Kota Bandar Lampung memiliki tingkat pengangguran tertinggi di provinsi tersebut.

Untuk mengatasi masalah ini, Pemerintah Provinsi Lampung meluncurkan aplikasi SiGajah Lampung, yang memudahkan pencari kerja dan perusahaan untuk mencari dan memberikan informasi lowongan kerja, serta menjembatani kebutuhan kualitas tenaga kerja melalui program pelatihan dan pemagangan. Pemerintah Kota Bandar Lampung juga berupaya memudahkan anak muda dan UMKM untuk membuka usaha dengan mempermudah perizinan secara online.

Dalam konteks ini, pemodelan dan analisis data menjadi penting dalam memahami dan mengatasi pengangguran. Dalam penelitian ini, algoritma K-Nearest Neighbors (KNN) dan Support Vector Machines (SVM) dibandingkan

ISSN: 2598-0256, E-ISSN: 2598-0238

dalam klasifikasi pengangguran di Provinsi Lampung. KNN adalah algoritma berbasis "tetangga terdekat" yang mempertimbangkan kesamaan dengan tetangga terdekat, sedangkan SVM menggunakan pendekatan pemisah linear atau kernel untuk mengklasifikasikan data.

Penelitian ini memiliki tujuan yang spesifik, antara lain menentukan faktor-faktor yang signifikan dalam klasifikasi pengangguran di Provinsi Lampung, membandingkan kinerja KNN dan SVM dalam klasifikasi pengangguran, mengidentifikasi algoritma yang paling efektif, dan memberikan rekomendasi kebijakan dan program penanggulangan pengangguran.

Dengan mencapai tujuan-tujuan tersebut, penelitian ini memberikan kontribusi dalam pemahaman faktor-faktor pengangguran di Provinsi Lampung, pemilihan metode klasifikasi yang efektif, rekomendasi kebijakan dan program penanggulangan pengangguran, serta pengembangan penelitian di bidang pemodelan data dan klasifikasi.

Penelitian ini juga memberikan manfaat praktis, seperti kontribusi dalam memahami faktor-faktor pengangguran, pemilihan metode klasifikasi yang efektif, dan rekomendasi kebijakan penanggulangan pengangguran. Selain itu, penelitian ini juga memiliki kontribusi dalam pengembangan penelitian lebih lanjut di bidang pemodelan data dan klasifikasi. Dengan demikian, penelitian ini memiliki manfaat yang penting dalam memahami dan mengatasi pengangguran di Provinsi Lampung.

2. KERANGKA TEORI

Pengangguran adalah suatu keadaan di mana individu yang termasuk dalam angkatan kerja tidak bekerja atau tidak dapat memperoleh pekerjaan. Faktor-faktor penyebab pengangguran meliputi ketidakseimbangan antara angkatan kerja dan kesempatan kerja, struktur lapangan kerja yang tidak seimbang, peranan dan aspirasi wanita dalam angkatan kerja, ketidakseimbangan penyediaan dan pemanfaatan tenaga kerja antar daerah, serta adanya guncangan ekonomi seperti pandemi.

Pengangguran memiliki dampak negatif terhadap perekonomian. Tingkat pengangguran yang tinggi dapat mengurangi pendapatan nasional, menyebabkan penurunan penerimaan pajak, dan menghambat kegiatan pembangunan. Dampak negatif juga dirasakan oleh individu dan masyarakat, seperti kehilangan mata pencaharian, kehilangan keterampilan, dan ketidakstabilan sosial-politik.

Algoritma K-Nearest Neighbor (K-NN) digunakan dalam klasifikasi data dengan mempertimbangkan k titik terdekat dan menetapkan tanda mayoritas. K-NN memiliki keunggulan sederhana, kuat dalam pencarian, dan efektif untuk data dalam skala kecil. Namun, algoritma ini membutuhkan penentuan nilai k yang optimal dan memiliki biaya komputasi yang tinggi.

Support Vector Machine (SVM) merupakan metode klasifikasi yang mencari fungsi pemisah (hyperplane) terbaik untuk memisahkan dua set data yang berbeda. SVM mencari margin hyperplane terbesar antara dua kelas. SVM memiliki performansi yang baik dan mampu menemukan solusi global optimal. Namun, SVM memerlukan tahapan training dan testing serta penentuan parameter seperti konstanta C dan fungsi kernel.

Klasifikasi adalah proses menilai objek data dan memasukkannya ke dalam kelas tertentu berdasarkan model yang telah disimpan dalam memori. Klasifikasi melibatkan pembangunan model untuk pengenalan, klasifikasi, atau prediksi pada data lain.

Dalam konteks pengangguran di Provinsi Lampung, akan dilakukan perbandingan kinerja antara algoritma K-NN dan SVM. Penelitian ini akan menganalisis faktor-faktor pengangguran, memilih nilai k optimal untuk K-NN, dan mencari hyperplane terbaik untuk SVM. Hasil penelitian ini akan memberikan pemahaman yang lebih baik tentang faktor-faktor pengangguran di Provinsi Lampung dan memilih metode klasifikasi yang efektif untuk mengatasi masalah tersebut.

Tabel 1. Tinjauan Penelitian Terdahulu

No.	Peneliti	Judul	Metode	Hasil
1.	Ida Ayu Ade Sita Pratiwia, Arie Wahyu Wijayanto (2019)	Klasifikasi Indeks Pembangunan Manusia dengan Metode K-*Nearest Neighbor* dan *Support Vector Machine* di Pulau Jawa(Ade et al., n.d.).	Pada penelitian ini metode yang digunakan adalah K-*Nearest Neighbor* dan *Support Vector Machine* yang bertujuan untuk membandingkan akurasi klasifikasi Indeks Pembangunan Manusia (IPM) kabupaten/kota di Pulau Jawa tahun 2019.	Perbandingan dua metode tersebut dengan hasil akurasi menunjukkan bahwa metode yang terbaik untuk pengklasifikasian Indeks Pembangunan Manusia (IPM) adalah metode *Support Vector Machine* (SVM). Selain menggunakan hasil akurasi, penelitian ini menggunakan kurva ROC yang dapat membantu dalam menentukan metode mana yang terbaik.
2.	F Fauzi (2017)	K-*Nearset Neighbor* (K-NN) dan *Support Vector Machine* (SVM) untuk Klasifikasi Indeks	Pada penelitian ini metode yang digunakan adalah metode K-*Nearset Neighbor* (K-NN) dan	Hasil klasifikasi Indeks Pembangunan Manusia (IPM) dengan metode K- *Nearset Neighbor* (K-NN) dan *Support Vector Machine* (SVM) termasuk dalam tingkat klasifikasi

ISSN: 2598-0256, E-ISSN: 2598-0238

		Pembangunan Manusia Provinsi Jawa Tengah. (Fauzi, 2017)	*Suppport Vector Machine* (SVM).	sangat baik. Kedua metode tersebut jika dibandigkan metode *Support Vector Machine* (SVM) dengan kombinasi parameter gamma 1 ?? = 1, 10, dan 100 merupakan metode yang paling tepat dibandingkan dengan metode k-*Nearset Neighbor* (k-NN).
3.	Adhitya Prayoga Permana, Kurniyatul Ainiyah, Khadijah Fahmi Hayati Holle (2021)	Analisis Perbandingan Algoritma *Decision Tree*, KNN, dan *Naïve Bayes* untuk Prediksi Kesuksesan Start-up(Prayoga Permana et al., 2021).	Pada penelitian ini metode yang digunakan adalah klasifikasi algoritma *Decision Tree, Naïve Bayes* dan KNN.	Berdasarkan hasil per bandingan diantara ke tiga algoritma tersebut algoritma *Decision Tree* merupakan algoritma yang paling cocok untuk dig unakan di antara algoritma KNN dan *Naïve Bayes*. Begitupun dengan nilai presisinya
4.	Viona Novalia, Rito Goejantoro, dan Sifriyani (2020)	Perbandingan Metode Klasifikasi *Naive Bayes* dan K-*Nearest Neighbor* (Studi Kasus : Status Kerja Penduduk Di Kabupaten Kutai Kartanegara Tahun 2018). (Novalia et al., 2020)	Pada penelitian ini metode yang digunakan adalah *Naive Bayes* and *k-nearest neighbor*.	Metode klasifikasi *naive Bayes* dan *k-nearest neighbor* dapat dipergunakan dalam mengklasifikasikan data status kerja penduduk. Namun, pengklasifikasian menggunakan metode k-*nearest neighbor* memiliki ketepatan kinerja yang lebih baik dibandingkan dengan metode *Naive Bayes* dalam mengklasifikasikan status kerja penduduk
5	Fauziah, Muhammad Arif Tiro, & Ruliana	Comparison of K-*Nearest Neighbor* (K-NN) and *Support Vector Machine* (SVM) *Methods for Classification of Poverty Data in Papua*(Fauziah et al., 2022).(Fauziah et al., 2022)	Pada penelitian ini metode yang digunakan adalah *Support Vector Machine* dan K-*Nearest Neighbor*.	Kedua metode tersebut jika dibandingkan metode *Support Vector Machine* dengan *Parameter cost* = 1 merupakan metode paling tepat dibandingkan dengan metode K-NN dalam data kemiskinan di Papua tahun 2019
6	Wella,NiMade Satvika Iswari, Ranny	Perbandingan Algoritma KNN, C4.5, dan *Naive Bayes* dalam Pengklasifikasian Kesegaran Ikan Meng gunakan Media Foto(Made Satvika Iswari, 2017).(Made Satvika Iswari, 2017)	Pada penelitian ini metode yang digunakan adalah KNN, C4.5, dan *Naïve Bayes*.	Berdasarkan hasil perbandingan dari ke tiga algoritma ini, KNN memiliki akurasi yang paling tinggi diantara algoritma lainnya. Dengan demikian algoritma KNN dinilai cocok untuk digunakan dalam klasifikasi kesegaran ikan berdasarkan citra digital ikan.
7	Marthin Luter Laia, Yudi Setyawan.	Perbandingan Hasil Klasifikasi Curah Hujan Menggunakan *Metode Support Vector machine* dan *Naïve Bayes Classifier*(Laia & Setyawan, 2020). (Laia & Setyawan, 2020)	Pada penelitian ini menggunakan metode *Support Vector Machine* dan *Naïve Bayes Classifier*	Berdasarkan hasil analisis klasifikasi didapatkan bahwa metode terbaik yaitu *Support Vector Machine.* Hal ini dibuktikan dengan tingkat akurasi sebesar 79,45 % lebih besar dari tingkat akurasi metode *Naïve Bayes Classifier* yaitu 65,75%.
8	M.R Adrian, M.P. Putra, M.H. Rafialdy, N.A. Rakhmawati	Perbandingan Metode Klasifikasi *Random Forest* dan *Support Vector Machine* pada Analisis Sentimen PSBB(Adrian et al., n.d.). (Adrian et al., n.d.)	Pada penelitian ini menggunakan metode *Random Forest* dan *Support Vector Machine*	Dari penelitian tersebut *Support Vector Machine* dianggap lebih baik karena mampu mengenali tweet dengan label “Positif”
9	Isman, Andani Ahmad, Abdul Latief	Perbandingan Metode KNN dan LBPH Pada Klasifikasi Daun Herbal(Isman et al., 2021).(Isman et al., 2021)	Pada penelitian ini menggunakan metode K-*Nearest Neighbor* dan *Local Binary Pattern Histogram*	Pada pengujian ini didapatkan hasil akurasi pada pengujian K *Nearest Neighbor* lebih tinggi dibandingkan dengan metode *Local Binary Pattern Histogram*
10	Hiya Nalatissifa, Windu Gata, Sri Diantika, Khoirun Nisa	Perbandingan Kinerja Algoritma klasifikasi *Naïve Bayes, Support Vector Machine* (SVM), dan *Random Forest* untuk Prediksi Ketidakhadiran di Tempat Kerja(Nalatissifa et al., 2021).	Pada penelitian ini menggunakan metode *Naïve Bayes, Support Vector Machine*, dan *Random Forest*	Pada hasil penelitian, algoritma *Random Forest* memperoleh nilai akurasi, presisi, dan *recall* yang paling tinggi dibandingkan dengan algoritma *Naïve Bayes* dan SVM.

ISSN: 2598-0256, E-ISSN: 2598-0238

3. METODOLOGI

Proses pemodelan adalah suatu cara untuk menciptakan representasi atau abstraksi dari sebuah sistem, konsep, fenomena, atau objek dalam bentuk model. Tahap-tahap yang dilakukan pada proses pemodelan dimulai dari Pre Processing, Training Model, hingga proses metode klasifikasi menggunakan algoritma KNN dan SVM yang dilanjutkan dengan tuning parameter dan evaluasi.

Pada tahap Pre Processing, dilakukan langkah-langkah untuk memastikan data siap dan sesuai untuk proses selanjutnya, serta meningkatkan kualitas data dan menghilangkan masalah atau hambatan yang mungkin muncul. Tahap ini mencakup Seleksi Fitur dan Eksploratory Data Analysis.

Untuk memilih subset fitur yang paling relevan dan berpengaruh terhadap pemodelan, perlu dilakukan analisis mendalam terhadap setiap fitur yang ada dalam dataset. Pada tahap Seleksi Fitur, dilakukan identifikasi fitur yang tidak relevan dan menghapusnya, serta mengidentifikasi data duplikat atau tidak lengkap.

Sedangkan pada tahap Eksploratory Data Analysis, dilakukan identifikasi dan penanganan nilai yang hilang (missing values) dalam dataset. Data dibagi menjadi input / fitur dan target / label, dan dilakukan transformasi pada data yang dilakukan normalisasi (scaling data ke rentang tertentu), standardisasi (mengubah data menjadi mean = 0 dan standard deviasi = 1) atau transformasi logaritmik.

Setelah Pre Processing dilakukan, dilanjutkan dengan tahap Training Model yang melatih model pembelajaran mesin atau model statistik menggunakan data pelatihan. Tahap ini bertujuan untuk menghasilkan model yang dapat melakukan prediksi atau analisis yang akurat berdasarkan pola-pola di dalam data.

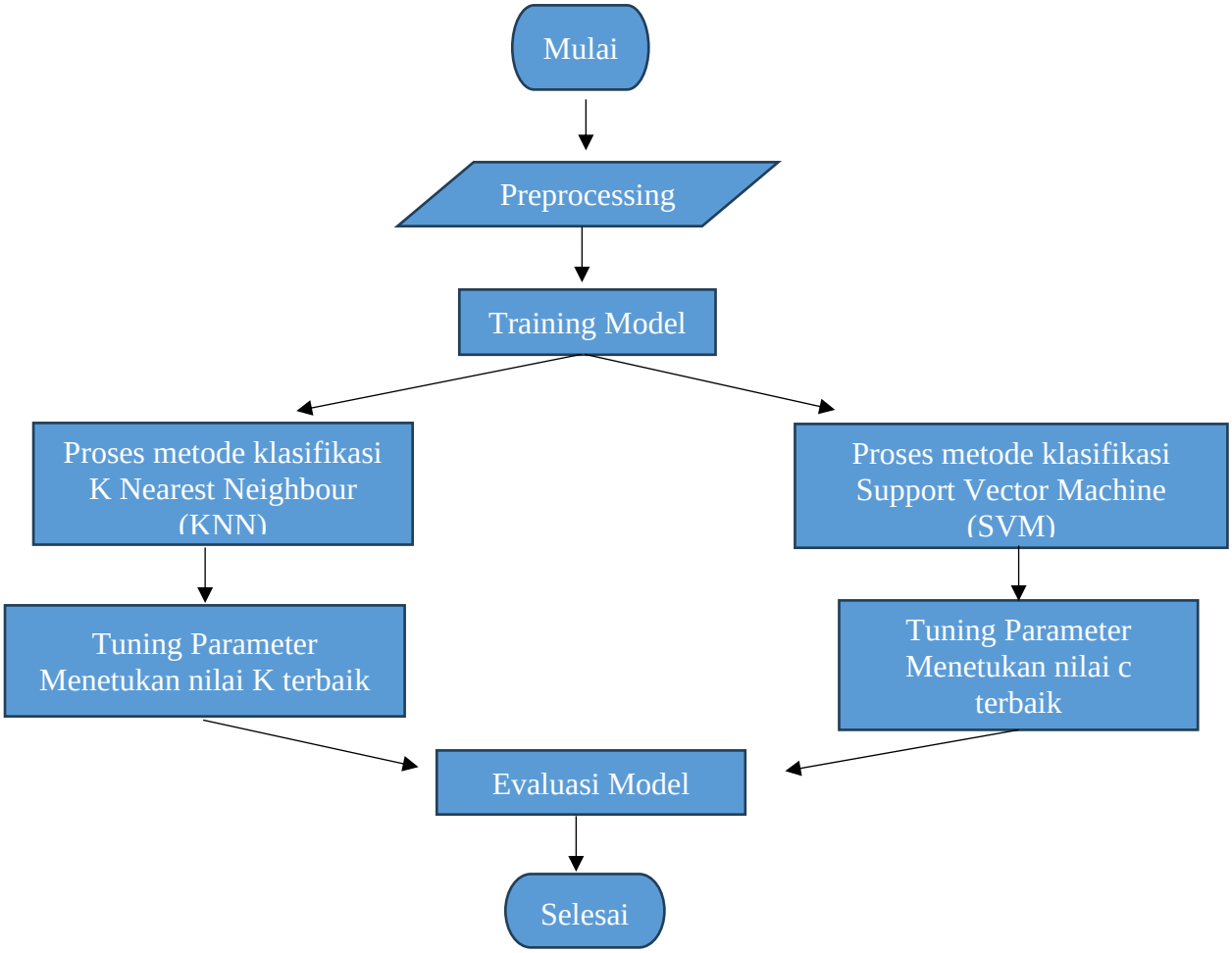
Tahap Training Model meliputi Train Val Test Split dan K-Fold Cross Validation. Train Val Test Split adalah proses membagi dataset menjadi tiga subset yang berbeda: data pelatihan (train set), data validasi (validation set), dan data pengujian (test set). Sementara itu K-Fold Cross Validation adalah sebuah metode yang digunakan untuk menguji dan mengevaluasi model statistik atau pembelajaran mesin.

4. HASIL DAN PEMBAHASAN

Penelitian ini masih di tahap desain model, oleh sebab itu tindak lanjut dalam penelitian ini akan melanjutkan penelitian sampai hasil evaluasi algoritma.

Berikut adalah desain modelnya :

ISSN: 2598-0256, E-ISSN: 2598-0238



5. KESIMPULAN

Pengangguran merupakan isu penting yang harus segera ditangani karena dampaknya akan sangat berimbas ke mana mana. Salah satu cara yang bisa dilakukan yaitu membuat pemodelan tentang pengangguran. Dalam kasus kali ini digunakan dua buah algoritma yaitu KNN dan SVM. Dari kedua algoritma ini akan dibandingkan mana yang lebih baik hasilnya untuk mendapatkan model pengangguran yang terbaik.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada semua pihak yang telah membantu penelitian ini, khususnya sumber data penelitian.

DAFTAR PUSTAKA

Ade, I. A., Pratiwi, S., & Wijayanto, A. W. (n.d.). *Klasifikasi Indeks Pembangunan Manusia dengan Metode K-Nearest Neighbor dan Support Vector Machine di Pulau Jawa*.

Adrian, M. R., Putra, M. P., Rafialdy, M. H., & Rakhmawati, N. A. (n.d.). *Perbandingan Metode Klasifikasi Random Forest dan SVM pada Analisis Sentimen PSBB, Adrian, Putra, Rafialdy, Rakhmawati*.

ISSN: 2598-0256, E-ISSN: 2598-0238

Fauzi, F. (2017). K-Nearset Neighbor (K-NN) dan Support Vector Machine (SVM) untuk Klasifikasi Indeks Pembangunan Manusia Provinsi Jawa Tengah Info Artikel. *Jurnal MIPA*, *40*(2). http://journal.unnes.ac.id/nju/index.php/JM

Fauziah, Tiro, M. A., & Ruliana. (2022). Comparison of k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) Methods for Classification of Poverty Data in Papua. *ARRUS Journal of Mathematics and Applied Science*, *2*(2), 83–91. https://doi.org/10.35877/mathscience741

Isman, Andani Ahmad, & Abdul Latief. (2021). Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(3), 557–564. https://doi.org/10.29207/resti.v5i3.3006

Laia, M. L., & Setyawan, Y. (2020). Perbandingan hasil klasifikasi curah hujan menggunakan metode SVM dan NBC, Marthin, Yudi. *Statistika Industri Dan Komputasi*, *05*, 51–61.

Made Satvika Iswari, N. (2017). Perbandingan Algoritma kNN, C4.5, dan Naive Bayes dalam Pengklasifikasian Kesegaran Ikan Menggunakan Media Foto. *114 ULTIMATICS*, *IX*(2). https://www.cs.waikato.ac.nz/ml/weka/

Nalatissifa, H., Gata, W., Diantika, S., & Nisa, K. (2021). Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja. *Jurnal Informatika Universitas Pamulang*, *5*(4), 578. https://doi.org/10.32493/informatika.v5i4.7575

Novalia, V., Goejantoro, R., & Sifriyani, D. (2020). Perbandingan Metode Klasifikasi Naive Bayes dan K-Nearest Neighbor (Studi Kasus : Status Kerja Penduduk Di Kabupaten Kutai Kartanegara Tahun 2018) The Comparison Method Of Classification Naive Bayes and K-Nearest Neighbor (Case Study: Employment Status Of Citizen In Kutai Kartanegara Regency 2018). *Jurnal EKSPONENSIAL*, *11*(2).

Prayoga Permana, A., Ainiyah, K., & Fahmi Hayati Holle, K. (2021). Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up. In *JISKa* (Vol. 6, Issue 3). https://www.kaggle.com/manishkc06/startup-success-prediction.