

Implementasi Metode Decision Tree Untuk Diagnosa Kehamilan

Proposal Proyek III

Oleh :

FAHIRA (1.20.40.44)

ZIAN ASTI DWIYANTI (1.20.40.49)



Universitas Logistik & Bisnis Internasional

Program Studi Diploma IV Teknik Informatika

Universitas Logistik dan Bisnis Internasional

Bandung

2022

Lembar Persetujuan

Implementasi Metode Decision Tree Untuk Diagnosa Kehamilan

Implementation of Decision Tree Method for Pregnancy Diagnosis

PROPOSAL PROYEK III

Program Studi DIV Teknik Informatika

Oleh :

FAHIRA (1.20.40.44)

ZIAN ASTI DWIYANTI (1.20.40.49)

Proposal ini diajukan sebagai usulan pembuatan proyek III pada
Program Studi Diploma IV Teknik Informatika
Universitas Logistik dan Bisnis Internasional

Bandung, 20 Oktober 2022

Menyetujui,

Pembimbing

Roni Habibi, S.Kom., M.T., SFPC

NIDN.0423127804

ABSTRAK

Menjadi ibu bukanlah hal yang mudah karena harus ada kehamilan dan persalinan yang ekstensif. Kehamilan adalah proses sembilan bulan atau lebih dimana harus memperhatikan kesehatan kandungan sedini mungkin. Bahkan hamil di usia muda lebih berisiko mengalami masalah kehamilan. Dalam bidang kesehatan telah dikembangkan suatu kecerdasan yang disebut machine learning yaitu suatu sistem komputer yang menggunakan pengetahuan, fakta dan metode untuk memecahkan masalah yang seringkali tidak dipahami oleh seorang ahli di bidangnya. Penelitian ini nantinya akan menggunakan metode Decision Tree yang akan melakukan pencarian mendalam berdasarkan pola data yang ada. Tujuan dari penelitian ini adalah untuk membantu ibu melakukan diagnosis kehamilan yang diprediksi normal selama kehamilan tanpa harus ke dokter. Data yang digunakan adalah usia, usia kehamilan, berat dan tinggi badan, sistole, diastole, IMT, lingkar perut dan variabel targetnya adalah diagnosis Normal atau tidak. Hasil dari penelitian ini akan menghasilkan suatu model yang akan diterapkan ke dalam sistem informasi klinik untuk mendiagnosa penyakit kehamilan.

Kata Kunci : Kehamilan, Decision Tree, Prediksi, Diagnosis

ABSTRACT

Becoming a mother is not an easy thing because there must be an extensive pregnancy and childbirth. Pregnancy is a process of nine months or more in which you should pay attention to the health of the uterus as early as possible. Even pregnant at a young age is more at risk of having pregnancy problems. In the health sector, intelligence called machine learning has been developed, which is a computer system that uses knowledge, facts and methods to solve problems that are often not understood by experts in their fields. This study will use the Decision Tree method which will perform an in-depth search based on existing data patterns. The purpose of this study is to help mothers make a pregnancy diagnosis that is predicted to be normal during pregnancy without having to see a doctor. The data used are age, gestational age, weight and height, systolic, diastolic, BMI, abdominal circumference and the target variable is a normal diagnosis or not. The results of this study will produce a model that will be applied to a clinical information system to diagnose pregnancy diseases

Keywords: Pregnancy, Decision Tree, Prediction, Diagnosis

A. Judul

Implementasi Metode Decision Tree Untuk Diagnosa Kehamilan

B. Deskripsi Aplikasi

Sistem informasi klinik yang sudah dibuat oleh penulis pada penelitian sebelumnya akan dikembangkan dengan menerapkan machine learning sebagai fitur prediksi diagnosa kehamilan untuk seorang ibu. Fitur prediksi disini menggunakan algoritma supervised learning dengan metode Decision Tree, outputnya akan menghasilkan model yang akan diterapkan ke sistem. Sistem yang sudah dikembangkan menjadi suatu fitur prediksi untuk diagnosa kehamilan.

C. Latar Belakang

Data adalah kumpulan sebuah informasi atau deskripsi tentang sesuatu yang diperoleh dengan mengamati atau mencari sumber tertentu. Data yang diperoleh dapat berupa asumsi atau fakta karena belum diolah lebih lanjut. Setelah diolah dengan penelitian atau eksperimen, data dapat berbentuk lebih kompleks seperti database, informasi atau bahkan solusi untuk menyelesaikan masalah tertentu. Data juga diperlukan untuk diolah dan disajikan dalam dunia bisnis.

Menurut data yang didapatkan Indonesia merupakan salah satu Negara dengan jumlah dan pertumbuhan penduduk yang besar. Menyadur dari kabar berita Kompas.com pada tanggal 03 Februari 2022 bahwa jumlah penduduk di Indonesia terus meningkat setiap tahunnya. Hasil sensus tahun 1980 berjumlah 147 juta jiwa dengan persentase pertumbuhan sebesar 2,34 persen per tahun. Keadaan pertumbuhan penduduk di Indonesia tentunya memerlukan perhatian yang besar untuk mengurangi jumlah kematian. Salah satu cara yang bisa dilakukan yaitu dengan memberikan kemudahan Ibu hamil untuk mengetahui kondisi kehamilannya.

Dalam era sekarang ini muncul sebuah bidang profesi baru yaitu data

science. Data science merupakan gabungan dari inferensi data, pengembangan algoritma dan juga teknologi untuk memecahkan masalah analitik yang kompleks. Dalam data science juga terdapat analisis prediktif suatu data untuk difilter dan ditemukan data yang benar agar menghasilkan suatu data yang akurat sesuai dengan data yang sebenarnya. Data science atau ilmu data adalah seperangkat prinsip fundamental yang mendukung dan memandu ekstraksi berprinsip informasi dan pengetahuan dari data. Ada ratusan jenis metode atau algoritma yang ada pada data science salah satunya yang ingin penulis jelaskan di proposal ini adalah klasifikasi.

Proses klasifikasi pada penelitian ini akan mengklasifikasikan diagnosa kehamilan, dimana variabel yang berpengaruh untuk saat ini penulis teliti adalah usia, usia kehamilan, berat dan tinggi badan, sistole, diastole, IMT, lingkar perut dan variabel targetnya adalah diagnosis normal atau tidak. Penelitian ini juga akan menggunakan Decision Tree karena metode tersebut sangat mudah diterapkan dalam memprediksi 2 label seperti yes atau no. Maka dari itu dalam proyek III ini kami mengambil judul “Implementasi Metode Decision Tree Untuk Diagnosa Kehamilan”.

D. Identifikasi Masalah

Dari beberapa uraian yang dikemukakan pada latar belakang, maka dapat diidentifikasi masalah-masalah sebagai berikut :

1. Bagaimana melakukan prediksi diagnosis kehamilan menggunakan klasifikasi ?
2. Bagaimana menerapkan klasifikasi menggunakan Decision Tree ?
3. Bagaimana cara seorang ibu bisa mengetahui diagnosis kehamilan tanpa harus ke dokter ?

E. Tujuan

1. Menerapkan fitur prediksi terkait diagnosis kehamilan menggunakan teknik machine learning.
2. Melakukan implementasi Decision Tree dengan bahasa Python dan

algoritma Supervised Learning.

3. Dengan membuat suatu sistem prediksi diagnosis kehamilan.

F. Lingkup Dokumentasi

1. Metode yang dipakai ialah Decision Tree.
2. Data yang digunakan adalah data yang menggunakan variabel usia, usia kehamilan, berat dan tinggi badan, sistole, diastole, IMT, lingkar perut dan variabel targetnya adalah diagnosis normal atau tidak.

G. Kajian Pustaka

a. Klinik

Klinik adalah salah satu sarana pelayanan kesehatan yang banyak dicari dan dibutuhkan dalam menunjang peningkatan kesehatan, baik secara individu maupun masyarakat. Klinik dituntut untuk memberikan pelayanan yang memadai dan memuaskan agar kebutuhan masyarakat dalam pelayanan kesehatan terpenuhi[1].

b. Sistem Informasi

Sistem informasi adalah suatu sistem didalam suatu organisasi yang mempertemukan kebutuhan pengelolaan transaksi harian, mendukung operasi, bersifat manajerial, dan kegiatan strategi dari suatu organisasi dan menyediakan pihak luar tertentu dengan laporan-laporan yang dibutuhkan(Japerson,2014) [2].

Pengertian sistem informasi adalah sistem yang menyediakan informasi dengan cara sedemikian rupa sehingga bermanfaat bagi penerima(Nurlalela, 2013). Secara lebih detail sistem informasi dapat didefinisikan sebagai seperangkat entitas yang terdiri dari hardware, software, dan brainware yang saling bekerjasama untuk menyediakan data yang diolah sehingga berguna dan bermanfaat bagi penerima data tersebut[3].

c. Website

Website atau situs dapat diartikan sebagai kumpulan halaman-halaman yang digunakan untuk menampilkan informasi teks, gambar diam atau gerak, animasi, suara, dan atau gabungan dari semuanya baik yang bersifat statis maupun dinamis yang membentuk satu rangkaian bangunan yang saling terkait, yang masing-masing dihubungkan dengan jaringan-jaringan halaman(Febrin Aulia,2015) [4].

Dimuat dalam jurnal Penda Sudarto (2018) bahwa menurut pendapat Arief (2011:7), “Web adalah salah satu aplikasi yang berisikan dokumen–dokumen multimedia (teks, gambar, suara, animasi, video) di dalamnya yang menggunakan protokol HTTP (hypertext transfer protocol) dan untuk mengakses menggunakan perangkat lunak yang disebut browser”[5].

d. Kehamilan

Kehamilan adalah serangkaian proses yang dimulai dengan pembuahan atau pertemuan sel telur yang sehat dengan cairan dan dilanjutkan dengan pembuahan, pembuahan, dan daya tarik. Masa kehamilan dibagi menjadi tiga perempat (280 hari 40 minggu atau 9 bulan 7 hari) [6].

e. Data Science

Data science merupakan gabungan dari inferensi data, pengembangan algoritma dan juga teknologi untuk memecahkan masalah analitik yang kompleks. Dalam data science juga terdapat analisis prediktif suatu data untuk difilter dan ditemukan data yang benar agar menghasilkan suatu data yang akurat sesuai dengan data yang sebenarnya. Data science atau ilmu data adalah seperangkat prinsip fundamental yang mendukung dan memandu ekstraksi berprinsip informasi dan pengetahuan dari data [7].

f. Klasifikasi

Algoritma klasifikasi memprediksi satu atau lebih variabel diskrit, berdasarkan atribut lain dalam dataset. Perangkat lunak data mining diperlukan untuk menjalankan algoritma klasifikasi. Metode klasifikasi merupakan salah satu metode statistika untuk mengelompok atau

mengklasifikasi suatu data yang disusun secara sistematis ke dalam suatu kelompok sehingga dapat diketahui suatu individu berada pada kelompok tertentu. Masalah klasifikasi ini muncul ketika terdapat sejumlah ukuran yang terdiri dari satu atau beberapa kategori yang tidak dapat didefinisikan secara langsung tetapi harus menggunakan suatu ukuran. Dengan seperti ini maka klasifikasi bisa melakukan prediksi untuk masalah diagnosa kehamilan [8].

g. Machine Learning

Machine Learning adalah sebuah istilah pembelajaran mesin yang diperkenalkan oleh Adrien Marie Legendre. Salah satu bidang terpenting dari kecerdasan buatan adalah pembelajaran mesin (AI). Tanpa perlu pemrograman khusus, mesin pembelajaran mesin dapat belajar dari pengalaman (atau lebih tepatnya) dengan cara yang sama seperti yang dilakukan manusia. Aplikasi ini dapat belajar, tumbuh, berubah, dan meningkatkan diri ketika diberikan informasi baru. Dengan kata lain, gagasan pembelajaran mesin adalah sistem komputer untuk menemukan pengetahuan yang berguna tanpa didikte ke arah mana harus mencari. Sebaliknya, mereka mencapai ini dengan menggunakan proses berulang, belajar dari data. Yang paling canggih, pembelajaran mesin adalah tentang kemampuan untuk beradaptasi dengan informasi baru secara mandiri dan berurutan [9].

h. Python

Python adalah bahasa pemrograman serbaguna untuk keperluan umum digunakan di banyak bidang seperti web, data, statistik, seluler. Ketika Bahasa R sering terbatas untuk bekerja di bidang analisis statistik. Dengan Python, kita bisa belajar dan membangun kepercayaan diri untuk menggunakannya konsep pemrograman umum dalam bahasa pemrograman populer lainnya seperti Java, C, atau C++. Python juga mengkompilasi program konsep umum dengan penekanan pada persiapan, operasi, dan pengiriman catatan. Selain itu, python adalah alat pembelajaran yang kuat, seperti pembelajaran mesin atau pembelajaran

mesin. Python telah menjadi alat untuk informasi ilmu pengetahuan atau ilmu informasi [10].

i. Decision Tree

Metode Decision Tree merupakan metode yang merubah fakta yang sangat besar menjadi sebuah pohon keputusan yang merepresentasikan aturan-aturan. Pohon keputusan ini juga berguna untuk mengeksplorasi data, serta menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, C4.5, CART. Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun [11].

H. Penelitian Terkait

No.	Area Penelitian	Tahun	Karakteristik Data	Metode	Model	Hasil Penelitian	Evaluasi (Akurasi)
1.	Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web[12].	2021	Data penyakit selama masa kehamilan dan gejalanya.	Machine Learning	Naive Bayes	Penerapan metode naive bayes untuk diagnosa penyakit selama kehamilan.	Ketepatan diagnosa yang diperoleh dari sistem yang sama dengan diagnosa dokter yaitu dengan persentase nilai 77%.
2.	Sistem Pakar Untuk Diagnosa Penyakit Kehamilan Menggunakan Metode	2016	Data gejala dan penyakit kehamilan serta data aturan yang sesuai dengan	Machine Learning	Dempster-shafer dan Decision tree	Sistem dapat menampilkan hasil diagnosa berupa	Akurasi sistem sebesar 76%.

	Dempster-Shafer Dan Decision Tree[13].		gejala dan penyakitnya.			nama penyakit	
3.	Sistem Prototype Klasifikasi Risiko Kehamilan Dengan Algoritma k-Nearest Neighbor (k-NN)[14].	2022	Data Kohort ibu hamil berdasarkan (KSPR).	Machine Learning	k-Nearest Neighbor(k-NN)	Prototype sistem mampu melakukan klasifikasi risiko kehamilan.	Perhitungan sesuai KSPR tingkat akurasi mencapai 80%.
4.	Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree[15].	2016	Data gejala penyakit hipertensi pada ibu hamil.	Machine Learning	Decision Tree	Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree	Akurasi = 92.6573% Error/Gagal = 7.3427%.
5.	Studi Implementasi Adaptive Neuro Fuzzy Inference System Untuk Menentukan Normalitas Kehamilan[16].	2015	Data informasi tentang periode kehamilan dan berbagai keluhan selama kehamilan terkait normalitas kehamilan dari literatur dan pakar.	Machine Learning	Adaptive Neuro Fuzzy Inference System	Metode ANFIS dapat diimplementasikan untuk menentukan normalitas kehamilan.	Akurasi menggunakan rumus $MAPE = 95\%$ mengenali normalitas kehamilan pada 40 data dan 90% dapat mengenali gejala kehamilan.

I. Metode Penelitian

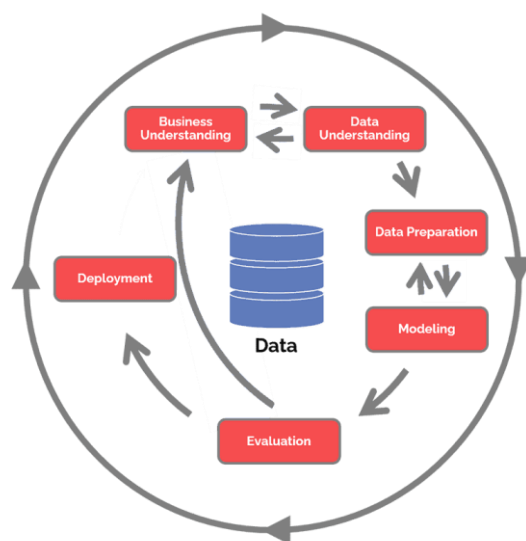
Metode yang digunakan dalam memecahkan masalah ini adalah analisis data secara kualitatif. Dengan metode analisis data secara kualitatif dan deskripsi dikembangkan atas dasar “kejadian” yang diperoleh ketika kegiatan

lapangan berlangsung. Karenanya, antara kegiatan pengumpulan data dan analisis data tidak mungkin dipisahkan satu sama lain. Sehingga kami menyelesaikan masalah dengan berdasarkan kejadian secara langsung dan proses bisnis dari sebuah klinik.

Metode pengumpulan data dilakukan dengan cara melakukan pengamatan langsung terhadap klinik dan menganalisis proses bisnis yang terjadi. Jenis penelitian yang digunakan adalah penelitian deskriptif analitik. Peneliti akan melakukan Analisa data yang terkumpul untuk mencari hubungannya dengan masalah yang terkait.

Untuk metode penelitian yang diterapkan pada penelitian ini menggunakan metode pengembangan waterfall. Metode waterfall ini merupakan metode pengembangan yang berurutan, dimulai dari tahap perencanaan hingga ke tahap pemeliharaan.

Adapun, selain menggunakan waterfall, penelitian ini juga akan menggunakan salah satu framework data science, agar mempermudah pengerjaan proses machine learning yang akan dibangun. Penulis mengambil referensi dengan mengikuti tahapan model Cross-Industry Standard Process for Data Mining (CRISP-DM). Adapun tahapan CRISP-DM, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment [17].



Tahapan – Tahapan CRISP-DM adalah :

1. Business Understanding

Business understanding adalah pemahaman terhadap tujuan bisnis, penilaian situasi dan menerjemahkan tujuan bisnis ke dalam tujuan data mining. Dalam penelitian ini membutuhkan pengetahuan untuk mendapatkan data dari beberapa sumber untuk diagnosa kehamilan dan tidak lupa pula melakukan analisa untuk membuat model prediksi tersebut.

2. Data Understanding

Pada tahap ini dilakukan proses pengumpulan data, kemudian melakukan analisa data serta melakukan evaluasi kualitas data yang digunakan dalam pengerjaan internship. Sumber data yang digunakan terdapat dari berbagai sumber dan penelitian terdahulu, serta melakukan interview ke beberapa expert. Maka sumber literatur banyak didapatkan dari buku, paper atau jurnal, karya ilmiah dan situs-situs penunjang.

3. Data Preparation

Data Preparation atau bisa disebut juga dengan data preprocessing adalah suatu proses/langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas (input yang baik untuk data mining tools).

4. Modelling

Dalam tahap ini, berbagai macam metode pemodelan dipilih dan diterapkan ke dataset yang sudah disiapkan untuk mengatasi kebutuhan bisnis yang sesuai. Adapun metode yang digunakan yaitu Decision Tree sebagai metode utama klasifikasi.

5. Evaluation Pada tahap terakhir ini, model yang sudah dibuat diuji dan dievaluasi keakuratan. Tahap ini mengukur sejauh mana model yang sudah dipilih memenuhi sasaran-sasaran bisnis dan bila demikian, sejauh manakah itu (apakah perlu lebih banyak model untuk dibuat dan diukur).

6. Deployment

Deployment adalah membuat sistem dari model yang telah dibuat dan

dikembangkan ke sistem yang telah ada sebelumnya. Pasti nya akan melihat evaluasi terlebih dahulu, jika evaluasi sudah baik, maka deployment dilakukan.

J. Jadwal Pelaksanaan

No	Kegiatan	Bulan							
		Oktober		November		Desember		Januari	
1	Tahap Perencanaan								
2	Tahap Pembuatan Aplikasi								
3	Tahap Pembuatan Laporan								
4	Tahap Implementasi								
5	Perbaikan								

*Keterangan: shading warna *grey*

DAFTAR PUSTAKA

- [1] Hutahaeen, Jeperson. 2014. *Konsep Sistem Informasi*. Yogyakarta: Deepublish.
- [2] Herlina, Asti, Prima Muhammad Rasyid. 2016. "Sistem Informasi Monitoring Pengembangan Software Pada Tahap Development Berbasis Web" dalam Jurnal Informatika, Volume III No.1.
- [3] Batubara, F. A. (2015). "Perancangan Website Pada PT. Ratu Enim Palembang". JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI TERAPAN" REINTEK"(REKAYASA INOVASI TEKNOLOGI), 7(1).
- [4] Hasugian, P. S. (2018). Perancangan Website Sebagai Media Promosi Dan Informasi. *Journal Of Informatic Pelita Nusantara*, 3(1).
- [5] SIANIPAR, Rismon Hasiholan. *Pemrograman Javascript: Teori Dan Implementasi*. Penerbit INFORMATIKA, 2015.
- [6] Handoko, M. R., & Neneng, N. (2021). Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web. *Jurnal Teknologi Dan Sistem Informasi*, 2(1), 50-58.
- [7] Ramirez-Figueroa, J. A., Martin-Barreiro, C., Nieto-Librero, A. B., Leiva, V., & Galindo-Villardón, M. P. (2021). A new principal component analysis by particle swarm optimization with an environmental application for data science. *Stochastic Environmental Research and Risk Assessment*, 35(10), 1969-1984.
- [8] Keleş, M. K. (2019). Breast cancer prediction and detection using data mining classification algorithms: a comparative study. *Tehnički vjesnik*, 26(1), 149-155.
- [9] Ahmad, M. A., Eckert, C., & Teredesai, A. (2018, August). Interpretable machine learning in healthcare. In *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics* (pp. 559-560).
- [10] Nehrey, M., & Hnot, T. (2021). Data Science Tools Application for Business Processes Modelling in Aviation. In *Research Anthology on Reliability and*

Safety in Aviation Systems, Spacecraft, and Air Transport (pp. 617-631). IGI Global.

- [11] Syarifah, S. J. P., Attaqwa, Q., & Pratama, A. (2022, September). Klasifikasi Menentukan Jadwal Kerja Data Karyawan Menggunakan Algoritma C4. 5 dan K-nearest Neighbor: Classification Determining Employee Data Work Schedules Using C4. 5 and K-Nearest Neighbor Algorithm. In SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat (Vol. 1, No. 1, pp. 215-221).
- [12] Handoko, M. R., & Neneng, N. (2021). Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web. *Jurnal Teknologi Dan Sistem Informasi*, 2(1), 50-58.
- [13] Minardi, J., & Suyatno. (2016). Sistem pakar Untuk Diagnosa Penyakit Kehamilan Menggunakan Metode Dempster-Shafer dan Decision Tree. *Jurnal SIMETRIS*, 7(1), 2252-4983.
- [14] A. Deharja, M. W. Santi, M. Yunus, and E. Rachmawati. (2022, Mei). “Sistem Prototype Klasifikasi Risiko Kehamilan Dengan Algoritma k-Nearest Neighbor (k-NN)”, *jtim*, vol. 4, no. 1, pp. 66-72.
- [15] Muzakir, A., & Wulandari, R. A. (2016, Mei). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1), p-ISSN 2407-7658.
- [16] Rusdiana, L., Sedyono, E., & Surarso, B. (2015). Studi Implementasi Adaptive Neuro Fuzzy Inference System Untuk Menentukan Normalitas kehamilan. *Jurnal Sistem Informasi Bisnis*.
- [17] Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526-534.