

Perbandingan Metode *K-Nearest Neighbor* dan Pohon Keputusan dalam Analisis Sentimen Data Ulasan Aplikasi Pinjaman *Online* Berizin OJK Di Google Play

Alya Firyal Laila Ramadhina¹, Erza Sofian²

¹²STIMIK ESQ; Cilandak Jakarta Selatan

e-mail: ¹a.firyal.l.r@students.esqbs.ac.id, ²erza.sofian@esqbs.ac.id

Diterima : 03 Januari 2024

Disetujui : 01 Februari 2024

Abstrak— Penelitian ini bertujuan untuk membandingkan nilai akurasi dari dua model algoritma klasifikasi *K-Nearest Neighbor* dan Pohon Keputusan dalam Data Ulasan Aplikasi Pinjaman Online Berizin OJK di Google Play yang mengandung nilai positif dan negatif dan membuktikan apakah semakin tinggi nilai akurasi model algoritma klasifikasi yang didapatkan menunjukkan semakin baik model algoritma klasifikasi tersebut dalam melakukan analisis sentimen. Metodologi penelitian yang digunakan CRISP-DM. Model algoritma klasifikasi yang digunakan untuk perbandingan sentimen adalah *K-Nearest Neighbor* dan Pohon Keputusan dengan menggunakan metode *Cross Validation* dan *Confusion Matrix* untuk pengujian model. Sampel dalam penelitian ini sebanyak 320 data ulasan pengguna dari beberapa aplikasi pinjaman online legal berizin OJK di Google Play pada kurun waktu 20 Januari 2023 sampai 31 Maret 2023. Hasil penelitian menunjukkan bahwa model algoritma *K-Nearest Neighbor* dengan memiliki tingkat akurasi yang lebih tinggi dengan tingkat akurasi sebesar 93.75% dan Pohon Keputusan dengan tingkat akurasi sebesar 76,56% sehingga dihasilkan nilai akurasi *K-Nearest Neighbor* lebih baik daripada nilai akurasi Pohon Keputusan dalam melakukan analisis sentimen. Kemudian hasil analisis sentimen terhadap data ulasan aplikasi pinjaman online berizin OJK di Google Play dengan jumlah sentimen negatif sebesar 63.7% dan sentimen positif sebesar 36.2%. Hasil penelitian ini diharapkan dapat membantu memberikan informasi terkait akurasi model algoritma klasifikasi yang baik dan mendapatkan hasil sentimen positif dan negatif dari beberapa layanan aplikasi Pinjaman online legal berizin OJK.

Keywords — Analisis Sentimen, Aplikasi Pinjol, CRISP-DM, Pohon Keputusan, *K-Nearest Neighbor*

I. PENDAHULUAN

Masyarakat Indonesia sempat dihebohkan dengan mudahnya meminjam uang melalui platform pinjaman online. Namun, maraknya pinjaman online juga memiliki dampak negatif yang perlu diwaspadai karena pada praktiknya menyisakan banyak problem di masyarakat yang tentunya sangat merugikan. Semakin banyak kasus pelanggaran yang dilakukan perusahaan pinjaman online ilegal atau tidak berizin. Pemerintah mengambil tindakan dengan memberantas dan menindak tegas praktik pinjaman online (pinjol) ilegal yang dapat merugikan masyarakat. Dan kini per 20 Januari 2023, data statistik Otoritas Jasa Keuangan (OJK) menyebutkan sebanyak 102 perusahaan menjadi

penyelenggara pinjol atau *fintech peerto-peer lending* berizin di Otoritas Jasa Keuangan (OJK) [7]. Maka dari itu pemerintah menghimbau masyarakat yang hendak melakukan pinjol untuk menggunakan platform pinjol yang legal atau berizin OJK saja.

Mengenai platform pinjol yang legal meskipun telah mendapatkan izin resmi, masih ada keluhan dari konsumen terkait layanan pinjol legal. Cara lain untuk mendapatkan umpan balik yang lebih mendalam dari masyarakat khususnya pengguna aplikasi adalah dengan memanfaatkan fitur ulasan (*review*) dari pengguna aplikasi pinjol legal di website Google Play Store. Ulasan tersebut mengandung opini dari para pengguna yang memiliki makna yang berbeda-beda sehingga

harus diselaraskan dengan memanfaatkan teknologi algoritma klasifikasi menggunakan model *machine learning* untuk melakukan analisis sentiment, sehingga dapat diketahui informasi tentang nada emosional ulasan tersebut positif atau negatif. Untuk menunjang penelitian tersebut diperlukan satu metode khusus untuk mengklasifikasi terhadap opini yang memiliki hasil akurasi yang lebih baik diantara dua metode *K-Nearest Neighbor* dan Pohon Keputusan yang dapat mengetahui ulasan konsumen terhadap aplikasi pinjol legal.

Adapun alasan penggunaan *K-Nearest Neighbor* dan Pohon Keputusan dalam penelitian. Di antara banyaknya metode klasifikasi, Pohon Keputusan merupakan algoritma yang kuat, populer, mudah ditafsirkan dan banyak diterapkan untuk beberapa masalah dalam data mining. Algoritma ini memberikan kinerja yang sangat baik dan mudah dimengerti [5]. Selain metode metode *K-Nearest Neighbor* dan metode pohon keputusan adalah model nonparametrik intuitif dan efektif yang digunakan untuk tujuan klasifikasi dan regresi. Karena keefektifannya, intuitif dan kesederhanaannya, *K-Nearest Neighbor* telah menarik minat luas dalam komunitas penelitian [4].

Dan alasan lain dari penggunaan metode *K-Nearest Neighbor* dan Pohon Keputusan dalam penelitian (Bayhaqy et al., 2018) berjudul *Sentiment analysis about e-commerce from tweets using decision tree, K-Nearest Neighbor, and Naïve Bayes* dimana hasil penelitian ini menunjukkan hasil akhir metode Pohon Keputusan memiliki klasifikasi dengan nilai akurasi yang lebih tinggi daripada metode *K-Nearest Neighbor* yaitu 80,00% banding 78,00% [2]. Namun dalam penelitian terdahulu (Pattiiha & Hendry, 2022) berjudul Perbandingan metode *k-nn, Naïve Bayes, Pohon Keputusan* untuk Analisis Sentimen tweet twitter Terkait Opini Terhadap PT Pal Indonesia, mendapatkan hasil yang sebaliknya dimana hasil penelitian ini menunjukkan metode *K-Nearest Neighbor* memiliki hasil akhir klasifikasi lebih tinggi daripada metode Pohon Keputusan yaitu 83,38% banding 81,09% [6]. Berdasarkan temuan diatas yang menunjukkan bahwa tidak ada machine

learning model yang secara konsisten mengungguli yang lain sehingga diputuskan untuk melakukan pengujian pada dua model klasifikasi pada penelitian ini.

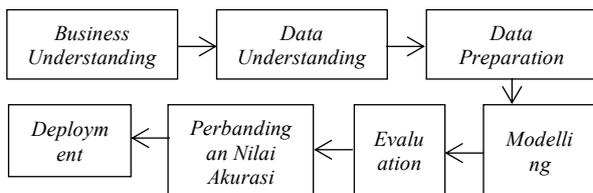
Untuk itu pada penelitian ini dilakukan komparasi antara dari dua model algoritma klasifikasi *K-Nearest Neighbor* dan Pohon Keputusan untuk klasifikasi ulasan konsumen aplikasi pinjol legal berizin OJK pada Januari 2023 sampai dengan Maret 2023. Hasil penelitian ini akan menjadi interpretasi mengenai aplikasi pinjol apakah opini konsumen yang berupa ulasan di google play lebih cenderung ke opini positif atau negatif dan membuktikan mengetahui apakah nilai akurasi *K-Nearest Neighbor* lebih baik daripada nilai akurasi Pohon Keputusan dalam melakukan analisis sentimen pada pada sekumpulan ulasan konsumen mengenai aplikasi pinjol legal berizin OJK sehingga model tersebut dapat direkomendasikan dalam melakukan analisis sentimen.

II. METODE PENELITIAN

Penelitian ini menggunakan metode studi *perbandingan (Comparative Study)* yang dilakukan dengan cara membandingkan dua metode algoritma dalam analisis sentimen untuk membandingkan tingkat akurasi dari kedua metode tersebut diantaranya adalah metode *K-Nearest Neighbor* dan metode Pohon keputusan. Metode yang digunakan dalam penelitian ini adalah penelitian kualitatif deskriptif analisis, yaitu data-data yang diperoleh kemudian ditampilkan dalam bentuk skema, lalu dideskripsikan sehingga dapat memberikan kejelasan informasi yang realistis. Dan data untuk penelitian ini diperoleh dari ulasan yang terdapat di website google play, yaitu data ulasan konsumen atau pengguna terhadap aplikasi-aplikasi pinjol legal.

Agar penelitian yang dilakukan mendapatkan hasil yang optimal yaitu untuk mengetahui model algoritma mana yang terbaik dalam analisis sentimen ini dengan melakukan perbandingan antara dua model algoritma. Maka pengerjaan penelitian dengan langkah-langkah pengerjaan terstruktur sangat diperlukan yaitu suatu metode yang dapat memperlancar jalannya pengerjaan

penelitian ini, metode itu adalah *CRISP-DM*, dengan rancangannya sebagai berikut,



Gambar 1 Rancangan Pemecahan Masalah

Dalam menganalisis sentimen, *CRISP-DM* memiliki enam tahapan agar hasil yang diperoleh dalam analisis optimal. Tahapan dalam *CRISP-DM* adalah Business understanding, Data understanding, Data preparation, Modelling, Evaluation, dan Deployment [3].

III. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Pada tahap ini dilakukan pemahaman terhadap objek penelitian. Salah satu permasalahan yang diangkat adalah adanya keluhan dari konsumen atau pengguna mengenai aplikasi jasa pinjol legal berizin OJK. Oleh karena itu, bentuk solusi yang ditawarkan adalah membuat sebuah model klasifikasi algoritma yang nantinya dapat digunakan untuk melakukan klasifikasi terhadap sentimen konsumen atau pengguna mengenai aplikasi pinjol legal tersebut. Pada tahap ini juga dilakukan pemahaman untuk mendapatkan metode klasifikasi yang lebih baik antara metode *K-Nearest Neighbor* dan metode *Pohon Keputusan* saat proses pengolahan data yang akan dilakukan dengan membandingkan hasil dari algoritma. Pengujian ini dilakukan juga untuk mendapatkan akurasi terbaik dalam proses pengujiannya

3.2 Data Understanding

Tahap ini adalah proses memahami data yang akan digunakan sebagai bahan yang akan diteliti untuk dapat dilanjutkan ke tahap selanjutnya yaitu Preprocessing. Mengumpulkan data ulasan pengguna dari beberapa aplikasi pinjol berizin OJK yang ada di google play store, dengan alamat domain dari situs web google play ini <https://play.google.com/store/apps/category/FINANCE?hl=en> dapat dilihat pada gambar

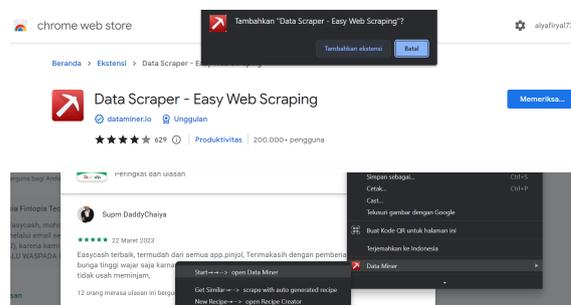


Gambar 2 Aplikasi Layanan Finance di PlayStore

Dipilih delapan aplikasi pinjol terpopuler di google play dari 20 aplikasi peringkat teratas bagian layanan keuangan (*finance*). Delapan aplikasi tersebut yaitu Easycash-Kredit Dana Online, Rupiah Cepat-Pinjaman Dana, Kredit Pintar-Pinjol, AdaKami-Pinjaman Uang Digital, Pinjamduit-KTA Cepat (OJK), UATAS-Pinjaman Dana Online, UangMe-Pinjaman Uang Online, dan BantuSaku-Pinjol OJK. Data yang dikumpulkan yaitu data ulasan pengguna pada rentang waktu 20 Januari 2023 – 31 Maret 2023

3.3 Data Preparation dan Preprocessing

Pengambilan data pada penelitian ini adalah dengan mengambil data ulasan pengguna dari beberapa aplikasi pinjol berizin OJK yang ada di google play store. Langkah awal yang dilakukan dalam persiapan data yaitu mengunduh aplikasi dan mengaktifkan ekstensi untuk scraping data yaitu Data Scraper ada pada Google Chrome. Pada tahap ini akan dilakukan scraping di halaman ulasan pengguna terhadap aplikasi dengan menggunakan ekstensi Data Scraper untuk pengambilan datanya.



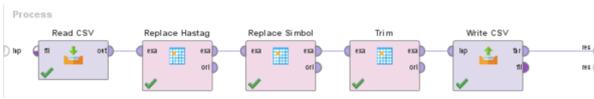
Gambar 3 Proses Scraping Data

Hasil dari scraping data yang telah dilakukan dapat disimpan dengan beberapa format yaitu CSV File dan Excel File. Ulasan yang digunakan hanya ulasan berbahasa Indonesia. Data yang didapatkan pada proses scraping data pada tahap sebelumnya belum dapat digunakan karena masih bentuk kalimat yang belum terstruktur dan belum

bersih sehingga perlu dilakukan proses *preprocessing* data. Tahap *preprocessing* ini dilakukan dengan menggunakan tools berupa Rapidminer. Adapun tahapan *preprocessing* yang akan dilakukan adalah sebagai berikut.

1. Pembersihan Data (*Cleansing Data*)

Pada bagian *Cleansing* ini dilakukan dengan tujuan untuk mengubah ataupun menghapus karakter atau kata tertentu yang ada pada data ulasan. Operator yang digunakan untuk pembersihan data pada penelitian ini yaitu *replace hashtag* dan *replace simbol* [!~?.,:!"#*%@%\$&()-]. Dapat dilihat penggunaan tools Rapid Miner dalam melakukan pembersihan data pada gambar 4 dan hasilnya pada tabel 2.



Gambar 4 Proses Cleansing Data

Tabel 1 Hasil Replace Hashtag dan Simbol

	Sebelum	Sesudah
<i>replace hashtag</i>	Tidak amanah... Pembayaran internet tidak sampai... #kapoksudah	Tidak amanah... Pembayaran internet tidak sampai... kapoksudah
<i>replace simbol/punctuation</i>	Proses mengisi datanya sangat mudah.. Semoga pengajuan saya di acc .. dan bisa memudahkan orang" yg sedang membutuhkan.. Saya kasih bintang lima ,,,, mudah2an cair supaya bisa membantu orang yg lagi membutuhkan ,,,,,,terima kasih banyak !!! ?	Proses mengisi datanya sangat mudah. Semoga pengajuan saya di acc dan bisa memudahkan orang yg sedang membutuhkan Saya kasih bintang lima mudah2an cair supaya bisa membantu orang yg lagi membutuhkan terima kasih banyak

Sumber: Diolah oleh peneliti (2023)

2. Pemberian label

Dalam penelitian ini data yang sudah terkumpul dan telah melewati tahap *Preprocessing*, maka selanjutnya merupakan tahap pelabelan. Pemberian label dilakukan secara manual. Pelabelan data mengacu kepada dictionary berkonotasi negatif dan positif. Proses pelabelan yang dilakukan secara manual berdasarkan pengetahuan domain individu dan pemahaman bahasa dimana prosesnya akan memakan waktu yang cukup banyak. Berikut contoh hasil

pemberian label sentimen pada setiap ulasan pada tabel 2.

Tabel 2 Pelabelan Dataset

Ulasan	Sentimen
Sangat buruk belum lewat masa jatuh tempo sudah di teror dengan katakata kasar dan mengancam ini sudah bisa di laporkan ke OJK dan badan pengawas sejenis lainnya Kapok mending uninstall Sampah	negatif
pinjol yang cukup membantu proses cepat mudah dan tidak banyak persyaratan apapun baru sekali mencoba Isung di kasih limit dan mengajukan Isung di acc utk yang mau menggunakan aplikasi apapun satu kuncinya klo mau di permudahkan ketika di beri kepercayaan jng pernah lalai dng kewajiban kita dlm membayar angsuran	positif

Dalam pemberian label pada suatu dataset diperlukan penambahan kolom baru dengan tujuan untuk memberi identifikasi terkatit dataset tersebut yang diberi nama kolom sentimen.

3. *Tokenizing*

Proses *Tokenizing* yaitu membagi sebuah kalimat menjadi beberapa bagian seperti kata per kata, Contoh dari *tokenizing* pada Tabel.

Tabel 3 Hasil Proses *Tokenizing*

Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
Mantab mudah dan cepat pencairannya dan bunganya kecil Saya akan tetap memakai aplikasi dan menjaga pembayaran tepat waktu	Mantab mudah dan cepat pencairannya dan bunganya kecil Saya akan tetap memakai aplikasi dan menjaga pembayaran tepat waktu

4. *Case Folding*

Pada tahap *preprocessing* terdapat bagian *Case folding* dengan tujuan untuk mengubah bentuk tulisan menjadi seragam seperti menjadi huruf kecil. Karena tidak semua dokumen teks memiliki penggunaan huruf kapital yang tepat.

Hal tersebut dilakukan agar kata yang memiliki huruf kecil dan besar tidak terdeteksi atau dianggap memiliki perbedaan arti atau makna. Pada penelitian ini dataset ulasan akan diseragamkan menjadi huruf kecil. Hasilnya sapat dilihat pada tabel 4 Hasil *Case Folding*.

Tabel 4 Hasil *Case Folding*

Sebelum <i>Case Folding</i>	Sesudah
Saya menggunakan aplikasi ini karena saya membutuhkannya untuk hal yang mendesak dan dengan jumlah yang tidak terlalu besar tidak MEREKOMENDASIKAN SAMA SEKALI DC PUSH ANDA MESKIPUN ANDA BELUM TERLAMBAT MENGGUNAKAN KATAKATA KERAS MENGANCAM DENGAN DATA DAN BARANG ANDA LALU TANGGAL WAKTU PEMBAYARAN SAYA BERUBAH	saya menggunakan aplikasi ini karena saya membutuhkannya untuk hal yang mendesak dan dengan jumlah yang tidak terlalu besar tidak merekomendasikan sama sekali dc push anda meskipun anda belum terlambat menggunakan katakata keras mengancam dengan data dan barang anda lalu tanggal waktu pembayaran saya berubah

5. Filter Token

Proses *filter tokens (by Length)* proses ini kata-kata yang memiliki panjang karakter kurang dari 4 dan lebih dari 25 akan dihapus, seperti kata di, ada, oleh yang merupakan kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen, hasilnya pada tabel 5.

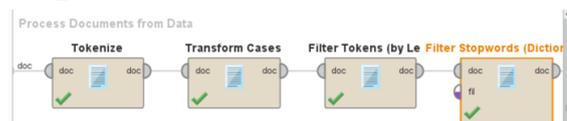
Tabel 5 Hasil dari Proses Filter tokens (by Length)

Sebelum <i>filter tokens</i>	Sesudah
mantab mudah dan cepat pencairannya dan bunganya kecil saya akan tetap memakai aplikasi dan menjaga pembayaran tepat waktu	mantab mudah cepat pencairannya bunganya kecil saya akan tetap memakai aplikasi menjaga pembayaran tepat waktu

6. Stopword Removal

1. Proses penghapusan kata-kata yang sering digunakan namun tidak memiliki makna yang penting menggunakan operator Stopwords removal. Contoh stopword dalam bahasa Indonesia adalah “yang”, “dari”, “di”, “dan”, “adalah”, dan sebagainya. Tujuan dari tahap ini adalah supaya dapat berfokus pada kata yang memiliki informasi penting sehingga dapat mencapai akurasi yang lebih baik dan akan mengurangi ukuran dataset dengan demikian akan mengurangi waktu proses pelatihan dikarenakan jumlah data yang digunakan akan berkurang. Pada penelitian ini menggunakan data stopword list yang sudah tersedia di Kaggle milik Oswin Rahadiyan Hartono (<https://www.kaggle.com/datasets/oswinrh/indonesian-stoplist?resource=download>) yaitu Indonesian Stoplist yang kemudian diberi nama file stopwordlistbahasa.csv. Adapun proses dari stopwords removal pada Gambar 5 dan hasilnya dapat dilihat pada tabel 6.

2.



Gambar 5 Proses *Stopword Removal*

Tabel 6 Hasil dari Proses *Stopword Removal*

Sebelum <i>Stopwords Removal</i>	Sesudah
mantab mudah cepat pencairannya bunganya kecil saya akan tetap memakai aplikasi menjaga pembayaran tepat waktu	mantab mudah cepat pencairannya bunganya memakai aplikasi menjaga pembayaran

7. Pembagian Data Latih dan Data Uji

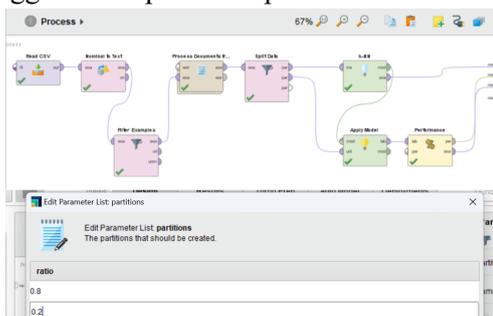
Setelah melakukan proses pelabelan data dari dataset yang berjumlah 320 selanjutnya adalah tahap pembagian data menjadi dua bagian, yaitu data latih (*data training*) dan data uji (*data testing*). Data latih ini nantinya digunakan untuk pembuatan model Machine Learning sedangkan data uji digunakan untuk menguji performa atau

kebenaran dari suatu model Machine Learning, dimana dalam penelitian ini model Machine Learning yang digunakan adalah klasifikasi.

$$JumlahDataLatih = Total\ Dataset \times \frac{ratio\ data\ latih}{100} \quad (1)$$

$$JumlahDataUji = Total\ Dataset \times \frac{ratio\ data\ uji}{100} \quad (2)$$

Berdasarkan jumlah sampel yang berhasil terkumpul yaitu 320 data ulasan aplikasi dilakukan pembagian dengan skenario 80:20 menggunakan cara manual didapat 256 sampel data ulasan untuk data latih (*data training*) dan 64 sampel data ulasan untuk data uji (*data testing*). Perhitungan tersebut sesuai jika pembagian data sampel tersebut dengan menggunakan fitur *Split Data* dengan *Sampling Type* yaitu *Stratified Sampling* pada aplikasi RapidMiner. *Stratified Sampling* adalah pemilihan tipe sampel dimana sampel dalam *dataset* dipilih secara acak tetapi tetap memastikan komposisi kelasnya atau *ratio* nya tetap. Berikut proses *split data* pada Gambar menggunakan aplikasi RapidMiner.



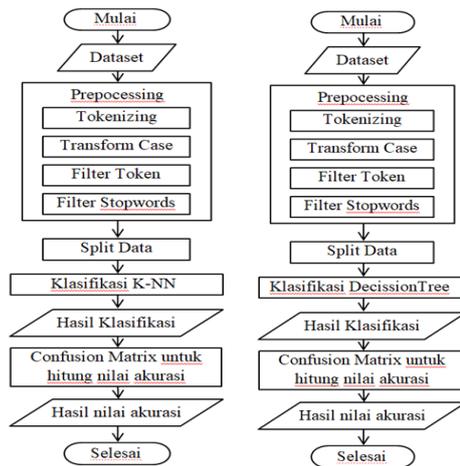
Gambar 6 Proses *Split Data*

Sampel data tersebut nantinya pada proses selanjutnya akan di terapkan kedalam model Machine Learning, yaitu *K-Nearest Neighbor* dan *Pohon Keputusan*.

3.4 Modeling

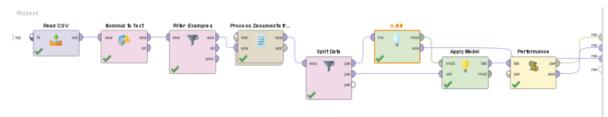
Setelah melakukan dan mendapatkan hasil *split data* adalah membuat classification model. Dalam pembuatan model, penelitian ini menggunakan alat bantu atau tools RapidMiner Studio versi 9.10. Dua model algoritma klasifikasi yang akan digunakan dalam penelitian ini adalah *K-Nearest Neighbor* dan *Pohon Keputusan*. Pembuatan model yang pertama adalah pembuatan model klasifikasi *K-Nearest*

Neighbor, yang kemudian dilanjutkan dengan model *Pohon Keputusan*.

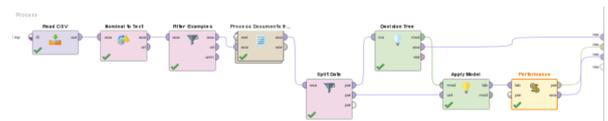


Gambar 7 Flowchart model klasifikasi *K-Nearest Neighbor* dan *Decision Tree*

Pembuatan model yang pertama adalah pembuatan model klasifikasi *K-Nearest Neighbor*. *Dataset* yang sudah bersih selanjutnya diterapkan pada model untuk mendapatkan hasil akurasi dari klasifikasi menggunakan algoritma model klasifikasi *K-Nearest Neighbor*. Berikut Gambar 8 proses penerapan pemodelan menggunakan model klasifikasi *K-Nearest Neighbor* pada aplikasi RapidMiner.



Gambar 8 proses pemodelan dengan *K-Nearest Neighbor*



Gambar 9 proses pemodelan dengan *Pohon Keputusan*

Split Data digunakan untuk mengambil *ExampleSet* sebagai inputnya dan mengirimkan subset dari *ExampleSet* tersebut melalui port outputnya. Jumlah subset (atau partisi) dan ukuran relatif dari setiap partisi ditentukan melalui parameter partisi. Jumlah rasio semua partisi harus 1 dimana ini digunakan untuk perbandingan pembagian jumlah *dataset*. Parameter tipe pengambilan sampel dengan menggunakan tipe *Stratified Sampling*. *Stratified Sampling* pemilihan tipe sampel dimana sampel

dalam *dataset* dipilih secara acak tetapi tetap memastikan komposisi kelasnya atau *ratio* nya tetap dan dalam penelitian ini perbandingan *ratio* nya 80:20. Langkah berikutnya data latih yang berjumlah 256 di latih menggunakan model K-NN dan Pohon Keputusan untuk dipelajari pola klasifikasi sentimen dan 64 data uji di aplikasikan pada K-NN dan Pohon Keputusan untuk mendapatkan tingkat akurasi dari masing-masing model klasifikasi.

Row No.	Sentimen	prediction%	confidence%	test	abukan	admin	apena	apah
1	Negatif	Negatif	1	0	0	0	0	0
2	Negatif	Negatif	1	0	0	0	0	0
3	Negatif	Negatif	0.998	0.402	0	0	0	0
4	Positif	Negatif	0.991	0.309	0	0	0	0
5	Negatif	Negatif	1	0	0	0	0	0
6	Negatif	Negatif	1	0	0	0	0	0
7	Negatif	Negatif	0.990	0.400	0	0	0	0
8	Negatif	Negatif	1	0	0	0	0	0
9	Negatif	Negatif	0.991	0.109	0	0	0	0
10	Negatif	Negatif	0.797	0.203	0	0	0	0
11	Negatif	Negatif	1	0	0	0	0	0
12	Positif	Positif	0.999	0.001	0	0	0	0

Gambar 10 Split Data Uji K-Nearest Neighbor

Row No.	Sentimen	prediction%	confidence%	test	abukan	admin	apena	apah
1	Negatif	Negatif	0.910	0.182	0	0	0	0.201
2	Negatif	Negatif	0.910	0.182	0	0	0	0
3	Negatif	Negatif	0.910	0.182	0	0	0.482	0
4	Positif	Positif	0	1	0	0	0	0
5	Negatif	Negatif	0.910	0.182	0	0	0	0
6	Negatif	Negatif	0.910	0.182	0	0	0	0
7	Negatif	Negatif	0.910	0.182	0	0	0	0
8	Negatif	Negatif	0.910	0.182	0	0	0	0
9	Negatif	Negatif	0.910	0.182	0	0	0	0
10	Negatif	Negatif	0.910	0.182	0	0	0	0
11	Negatif	Negatif	0.910	0.182	0	0	0	0
12	Positif	Positif	0.999	0.001	0	0	0	0

Gambar 11 Split Data Uji Pohon keputusan

Berikut adalah hasil *confusion matrix* model algoritma klasifikasi *K-Nearest Neighbor* pada tabel 7 dan Pohon Keputusan pada Tabel 8 di bawah ini.

Tabel 7 Hasil *Confussion Matrix K-Nearest Neighbor*

	True Negatif	True Positif
Prediksi Negatif	38(TN)	2(FN)
Prediksi Positif	2(FP)	22(TP)

Tabel 8 Hasil *Confussion Matrix Pohon Keputusan*

	True Negatif	True Positif
Prediksi Negatif	37(TN)	12(FN)
Prediksi Positif	3(FP)	12(TP)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{22 + 38}{22 + 38 + 2 + 2} \times 100\% = 93.75\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\% = \frac{22}{22 + 2} \times 100\% = 91.67\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\% = \frac{22}{22 + 2} \times 100\% = 91.67\%$$

Jadi pada pengujian menggunakan model K-Nearest Neighbor didapatkan akurasi yang diperoleh dengan menggunakan 320 data yang terdiri dari 256 data latih dan 64 data uji menghasilkan akurasi 93.75%, *recall* 91.67%, dan *Presisi* 91.67%.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{12 + 37}{12 + 37 + 3 + 12} \times 100\% = 76.56\%$$

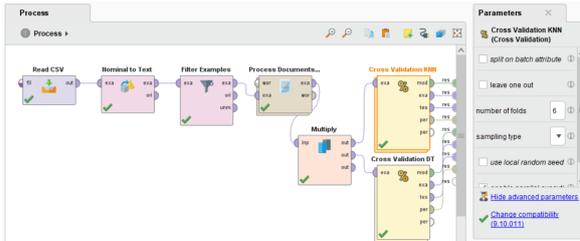
$$Recall = \frac{TP}{TP + FN} \times 100\% = \frac{12}{12 + 12} \times 100\% = 50\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\% = \frac{12}{22 + 3} \times 100\% = 80\%$$

Jadi pada pengujian akurasi yang diperoleh dengan menggunakan 320 data yang terdiri dari 256 data latih dan 64 data uji menghasilkan akurasi 76.56%, *recall* 50%, dan *Presisi* 80%.

3.5 Evaluation

Kemudian pada tahap ini dilakukan proses evaluasi performa pada kedua model *K-Nearest Neighbor* dan Pohon Keputusan dengan membandingkan kedua model tersebut menggunakan teknik *K-Fold Cross Validation* untuk mendapatkan hasil akurasi model yang lebih akurat dan lebih baik untuk mengklasifikasi sentimen. Pada Gambar 3. merupakan tahap penerapan operator *Cross Validation* pada kedua model klasifikasi dimana *number of fold* atau banyaknya iterasi pengujian dari *k = 2* sampai *k = 10* dengan *samplingtype* atau tipe samplingnya adalah *stratified sampling*



Gambar 12 Penerapan Cross Validation pada Model K-NN dan DT

Berikut adalah hasil perbandingan akurasi model algoritma klasifikasi *K-Nearest Neighbor* dan *Pohon Keputusan* menggunakan teknik *K-Fold Cross Validation* dengan jumlah iterasi pengujian hingga $k = 10$.

Tabel 9 Hasil Iterasi *Cross Validation*

	<i>K-Nearest Neighbor</i>		<i>Decision Tree</i>	
	<i>Accuracy</i>	<i>AUC</i>	<i>Accuracy</i>	<i>AUC</i>
$k = 1$	93.75%	0.982	76.56%	0.500
$k = 2$	85.94%	0.931	75.62%	0.500
$k = 3$	86.88%	0.936	71.88%	0.549
$k = 4$	88.13%	0.936	74.06%	0.595
$k = 5$	87.81%	0.936	74.38%	0.566
$k = 6$	88.16%	0.947	70.60%	0.527
$k = 7$	87.80%	0.940	74.38%	0.585
$k = 8$	87.81%	0.928	72.19%	0.625
$k = 9$	88.15%	0.930	75.64%	0.582
$k = 10$	87.50%	0.934	74.06%	0.568

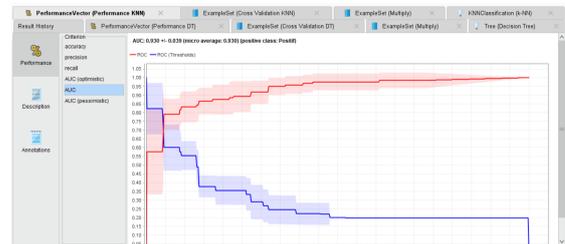
Berdasarkan hasil pengujian *Cross Validation* didapatkan untuk tingkat akurasi tertinggi pada *K-Nearest Neighbor* berada pada pengujian iterasi pertama atau $k = 1$ dengan tingkat akurasi 93.75% dengan AUC sebesar 98.2%. Kemudian untuk tingkat akurasi tertinggi pada *Pohon Keputusan* berada pada pengujian iterasi pertama atau $k = 1$ dengan hasil tingkat akurasi yaitu 76.56% dengan AUC sebesar 62.50% pada iterasi ke $k=8$. Sedangkan *K-Nearest Neighbor* pada pengujian iterasi ke $k = 2$ memiliki akurasi terendah yaitu 85.94% dan untuk *Pohon Keputusan* pada pengujian iterasi ke $k = 6$ memiliki akurasi terendah yaitu 70.60%.

3.6 Perbandingan

Berdasarkan hasil dari proses evaluasi menggunakan teknik *Cross Validation* didapatkan tabel perbandingan dibawah ini.

Tabel 10 Hasil Perbandingan Akurasi, AUC, Presisi, dan Recall

	Akurasi	AUC	Presisi	Recall
<i>K-Nearest Neighbor</i>	93.75%	98.20%	91.67 %	91.67%
<i>Decision Tree</i>	76.56%	50.00%	80.00%	50.00%



Gambar 13 Grafik AUC K-Nearest Neighbor



Gambar 14 Grafik AUC Decision Tree

Berdasarkan hasil perbandingan akurasi, AUC, Presisi, dan Recall *K-Nearest Neighbor* dan *Pohon Keputusan* pada Tabel 10 didapatkan bahwa hasil akurasi model *K-Nearest Neighbor* dengan tingkat akurasi 93.75% dengan AUC sebesar 98.2% (*Excellent Classification*). Dan hasil *Pohon Keputusan* mendapatkan akurasi 76.56% dengan AUC 50% (*failure Classification*) grafiknya dapat dilihat pada gambar 13 dan 14.

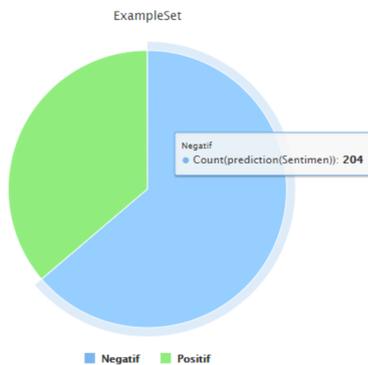
Telah dibuktikan bahwa model algoritma klasifikasi *K-Nearest Neighbor* memiliki hasil akurasi yang lebih tinggi dibandingkan *Pohon Keputusan* baik pada tahap pemodelan dan tahap evaluasi dimana penelitian ini memiliki hasil yang sejalan dengan beberapa penelitian sebelumnya terkait perbandingan metode algoritma klasifikasi *K-Nearest Neighbor* dan *Pohon Keputusan* seperti penelitian dari (Pattihha & Hendry, 2022) dengan menggunakan 1138 data mendapatkan tingkat akurasi pada metode *K-Nearest Neighbor* sebesar 83.38%, presisi 83.05% dan recall 99.89%. Serta tingkat akurasi pada *Pohon Keputusan* sebesar 81.09%, presisi 81.09% dan recall 100% [6]. Selanjutnya pada penelitian (Bangyal, W.H. et al., 2021) dengan menggunakan 10202 data mendapatkan tingkat akurasi pada metode *K-*

Nearest Neighbor sebesar 97% dan tingkat akurasi pada Pohon Keputusan sebesar 96% [2]. Kemudian pada penelitian (Tanggu Mara et al., 2021) dengan menggunakan 200 data mendapatkan tingkat akurasi pada metode *K-Nearest Neighbor* sebesar 87% dan nilai AUC sebesar 0.916 [8].

Kemudian untuk hasil dari tingkat akurasi kedua metode, pada penelitian ini selesai tingkat akurasi dari kedua metode cukup besar sedangkan di penelitian terdahulu selisihnya sebagian besar di bawah 5%. Perbedaan itu diduga disebabkan oleh keterbatasan penelitian pada cara pembersihan data yang dilakukan, pada penelitian sebelumnya yang mendapatkan selisih tingkat akurasi dibawah 5% melakukan proses pembersihan data yang lebih dalam dengan proses steaming untuk mengubah kata imbuhan menjadi bentuk dasarnya, sehingga hal ini memungkinkan untuk hasil akurasi yang didapatkan lebih signifikan. Hal itu yang diduga menjadi penyebab perbedaan antara penelitian terdahulu dengan penelitian ini dalam mengklasifikasi analisis sentimen pada aplikasi pinjol legal berizin OJK.

3.7 Deployment

Pada tahap ini, model yang lebih baik dijalankan berdasarkan tahap evaluasi sebelumnya guna mengetahui hasil analisis sentimen berupa jumlah sentimen positif dan jumlah sentimen negatif.

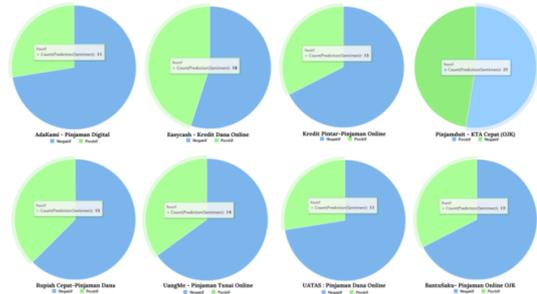


Gambar 15 Hasil Analisis sentimen dengan K-Nearest Neighbor

Tabel 11 Hasil Analisis sentimen dengan K-Nearest Neighbor

Index	Nominal value	Absolute count	Fraction
1	Negatif	204	0.637
2	Positif	116	0.632

Hasil analisis sentimen dari sekumpulan ulasan pengguna terkait aplikasi pinjaman hukum berizin OJK dengan total 320 data yang digunakan. Diagram berupa visualisasi *pie chart* pada Gambar menunjukkan sebanyak 63,70% pengguna memberikan ulasan negatif dan diikuti 36,20% merupakan ulasan positif terhadap aplikasi pinjaman online berizin OJK sejak 20 Januari 2023 hingga 31 Maret 2023.



Gambar 16 Hasil Sentimen Positif dan Negatif per Aplikasi dengan metode K-NN

Dari hasil tersebut dapat diketahui jumlah setiap sentiment positif dan negatif dari setiap aplikasi. Disini akan diurutkan aplikasi Pinjol yang memiliki paling banyak sentiment positif ke yang paling sedikit, sebagai berikut:

Pinjamduit-KTA Cepat (OJK) (Sentimen Positif=21), Easycash-Kredit Dana Online (Positif=18), Rupiah Cepat-Pinjaman Dana (Positif=15), UangMe-Pinjaman Uang Online (Positif=14), Kredit Pintar-Pinjol (Positif=13), BantuSaku-Pinjol OJK (Positif=13), AdaKami-Pinjaman Uang Digital (Positif=11), UATAS-Pinjaman Dana Online (ositif=11).

Berikut visualisasi *WordCloud* untuk kemunculan kata pada kelas positif dan negatif.



Gambar 17 Wordcloud Kelas Positif dan Kemunculan kata di Kelas Positif



Gambar 18 Wordcloud Kelas Positif dan Kemunculan kata di Kelas Negatif

Berdasarkan hasil visualisasi untuk kelas positif didapatkan banyak kata yang termasuk dalam kelas sentimen positif terhadap ulasan pelanggan pada aplikasi Pinjol legal di Google Play. Dan dapat diketahui kata-kata yang dapat mempengaruhi prediksi sentimen positif dari kata yang paling sering muncul. Begitu juga untuk kata-kata yang dapat mempengaruhi prediksi sentimen negatif.

IV. KESIMPULAN

Kesimpulan yang dapat diambil setelah melakukan pengolahan data adalah sebagai berikut:

1. Hasil *accuracy* model algoritma *K-Nearest Neighbor* lebih unggul dibandingkan model algoritma Pohon Keputusan dengan hasil rata-rata akurasi model *K-Nearest Neighbor* sebesar 93.75% dan Pohon Keputusan sebesar 76.56%. Sehingga hasil perbandingan pada penelitian ini terbukti bahwa nilai akurasi *K-Nearest Neighbor* lebih baik daripada nilai akurasi Pohon Keputusan dalam melakukan analisis sentimen pada pada sekumpulan ulasan pelanggan pada aplikasi Pinjol legal berizin OJK di Google Play.
2. Hasil analisis klasifikasi sentimen terhadap 320 data ulasan pelanggan pada aplikasi Pinjol legal menunjukkan kecenderungan pengguna merespon dengan ulasan negatif dengan jumlah sentimen negatif sebesar 63.70% dan sentimen positif sebesar 36.20%.

Adapun keterbatasan dan kekurangan dalam penelitian ini yaitu peneliti hanya mengolah data sejumlah 320 ulasan dari delapan aplikasi yang masing-masing 40 ulasan per-aplikasi. Dari penelitian yang telah dilakukan. Bentuk saran yang dapat diberikan pada peneliti selanjutnya antara lain:

1. Pengambilan data dapat dilakukan dengan jumlah data ulasan yang lebih banyak dengan tujuan menunjang keakuratan hasil.
2. Dapat melakukan proses pembersihan data dengan lebih maksimal dengan tujuan menunjang keakuratan hasil yang signifikan.

3. Dapat menggunakan model algoritma klasifikasi lainnya dalam menganalisis sentimen dengan tujuan untuk dapat menguji model dengan hasil yang lebih bervariasi.

Untuk mengurangi tingkat subjektivitas pada pelabelan sentimen salah satunya dapat menggunakan bantuan ahli bahasa.

DAFTAR PUSTAKA

- [1] Bangyal, W.H. *et al.* (2021) 'Detection of fake news text classification on covid-19 using Deep Learning Approaches', *Computational and Mathematical Methods in Medicine*, 2021, pp. 1–14. doi:10.1155/2021/5514220.
- [2] Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburuan, E. R. (2018). Sentiment analysis about e-commerce from tweets using decision tree, K-nearest neighbor, and Naïve Bayes. *2018 International Conference on Orange Technologies (ICOT)*. <https://doi.org/10.1109/icot.2018.8705796>
- [3] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. USA: SPSS Inc
- [4] Gou, J., Zhan, Y., Rao, Y., Shen, X., Wang, X., & He, W. (2014). Improved pseudo nearest neighbor classification. *Knowledge-Based Systems*, 70, 361–375. <https://doi.org/10.1016/j.knosys.2014.07.020>
- [5] Lakshmi, B. N., Indumathi, T. S., & Ravi, N. (2016). A study on C. 5Pohon Keputusanclassification algorithm for risk predictions during pregnancy. *Procedia Technology*, 24, 1542-1549.
- [6] Pattiiha, F. S., & Hendry, H. (2022). Perbandingan metode k-nn, Naïve Bayes, Pohon Keputusan Untuk Analisis Sentimen tweet twitter Terkait Opini TERHADAP PT Pal Indonesia. *JURIKOM (Jurnal Riset Komputer)*, 9(2), 506. <https://doi.org/10.30865/jurikom.v9i2.4016>
- [7] Syahputra, E. (2023, February 4). Daftar Terbaru Pinjol Legal 2023, Terdaftar & Berizin OJK. CNBC Indonesia. Retrieved April 11, 2023, from <https://www.cnbcindonesia.com/tech/20230204161012-37-410998/daftar-terbaru-pinjol-legal-2023-terdaftar-berizin-ojk>
- [8] Tanggu Mara, A., Sediyo, E. and Purnomo, H. (2021) 'Penerapan algoritma K-nearest neighbors Pada analisis sentimen metode Pembelajaran Dalam Jaringan (daring) Di Universitas Kristen Wira Wacana Sumba', *Jointer - Journal of Informatics Engineering*, 2(01), pp. 24–31. doi:10.53682/jointer.v2i01.30.