

Prediksi Harga Minyak Kelapa Sawit Dalam Investasi Dengan Membandingkan Algoritma Naïve Bayes, Support Vector Machine dan K-Nearest Neighbor

Deny Haryadi
President University
Bekasi, Indonesia
deniharyadi@gmail.com

Rila Mandala
President University
Bekasi, Indonesia
rilamandala@president.ac.id

Abstrak - Harga minyak kelapa sawit bisa mengalami kenaikan, penurunan maupun tetap setiap hari karena faktor yang mempengaruhi harga minyak kelapa sawit seperti harga minyak nabati lain (minyak kedelai dan minyak canola), harga minyak mentah dunia, maupun nilai tukar riil antara kurs dolar terhadap mata uang negara produsen (rupiah, ringgit, dan canada) atau mata uang negara konsumen (rupee). Untuk itu dibutuhkan prediksi harga minyak kelapa sawit yang cukup akurat agar para investor bisa mendapatkan keuntungan sesuai perencanaan yang dibuat. Tujuan dari penelitian ini yaitu untuk mengetahui perbandingan *accuracy*, *precision*, dan *recall* yang dihasilkan oleh algoritma Naïve Bayes, Support Vector Machine, dan K-Nearest Neighbor dalam menyelesaikan masalah prediksi harga minyak kelapa sawit dalam investasi. Berdasarkan hasil pengujian dalam penelitian yang telah dilakukan, algoritma Support Vector Machine memiliki *accuracy*, *precision*, dan *recall* dengan jumlah paling tinggi dibandingkan dengan algoritma Naïve Bayes dan algoritma K-Nearest Neighbor. Nilai *accuracy* tertinggi pada penelitian ini yaitu 82,46% dengan *precision* tertinggi yaitu 86% dan *recall* tertinggi yaitu 89,06%.

Keywords: Data Mining, Naïve Bayes, Support Vector Machine, K-Nearest Neighbor

I. PENDAHULUAN

Minyak kelapa sawit atau *Crude Palm Oil* (CPO) merupakan salah satu produk perkebunan yang menjadi salah satu komoditas utama ekspor Indonesia. Hal ini dikarenakan kondisi geografis wilayah Indonesia memang sangat cocok untuk pengembangan perkebunan kelapa sawit. Kebutuhan akan minyak kelapa sawit terus meningkat seiring peningkatan konsumsi minyak nabati dunia. Bahkan minyak sawit telah dikembangkan menjadi salah satu bahan bakar nabati untuk mengurangi dampak pemanasan global. Prospek pasar bagi minyak kelapa sawit cukup menjanjikan karena permintaan dari tahun ke tahun mengalami peningkatan yang cukup besar. Faktor utama pendorong kenaikan permintaan minyak kelapa sawit adalah harga yang relatif rendah dibandingkan dengan harga kompetitornya seperti minyak

kedelai dan minyak canola. Hal tersebut yang melatarbelakangi bahwa investasi minyak kelapa sawit sangat menjanjikan.

Investasi merupakan bentuk pengelolaan dana guna memberi keuntungan dengan cara menempatkan dana pada alokasi yang diperkirakan akan memberikan tambahan keuntungan. Umumnya investasi dibedakan menjadi dua yaitu investasi pada aset-aset finansial (*financial assets*) dan investasi pada aset-aset riil (*real assets*). Investasi pada aset-aset finansial dilakukan di pasar uang, misalnya berupa sertifikat deposito, *commercial paper*, surat berharga pasar uang, dan yang lainnya. Investasi juga dapat dilakukan di pasar modal, misalnya berupa saham, obligasi, waran, opsi, dan lain-lain. Sedangkan investasi pada aset-aset riil dapat berbentuk pendirian aset produktif, pendirian pabrik, pembukaan pertambangan, pembukaan perkebunan dan yang lainnya. Investasi pada minyak kelapa sawit digolongkan pada investasi riil (dalam bentuk yang bisa dilihat secara fisik). Proses berinvestasi ini memiliki tahapan yang penting yaitu melakukan analisis dengan mengidentifikasi perkiraan harga, kapan saat melepas dan berapa lama menahan kontrak sampai pada penyusunan kontrak jual-beli ke depan.

Harga minyak kelapa sawit bisa mengalami kenaikan, penurunan maupun tetap setiap hari karena faktor yang mempengaruhi harga minyak kelapa sawit seperti harga minyak nabati lain (minyak kedelai dan minyak canola), harga minyak mentah dunia, maupun nilai tukar riil antara kurs dolar terhadap mata uang negara produsen (rupiah, ringgit, dan canada) atau mata uang negara konsumen (rupee). Untuk itu dibutuhkan prediksi harga minyak kelapa sawit yang cukup akurat agar para investor bisa mendapatkan keuntungan sesuai perencanaan yang dibuat. Dalam hal ini peneliti membandingkan algoritma Naïve Bayes, Support Vector Machine dan K-Nearest Neighbor pada prediksi harga minyak kelapa sawit dalam investasi. Teorema Bayes merupakan teknik prediksi berdasarkan kemungkinan sederhana pada penerapan aturan Bayes dengan ketidaktergantungan yang kuat. Naïve Bayes banyak digunakan untuk proses klasifikasi karena Naïve Bayes lebih disukai dari kecepatan dan kesederhanaannya. Support Vector Machine merupakan algoritma yang memiliki keunggulan dalam optimasi sistem pengenalan pola dengan

kemampuan generalisasi yang baik. Support Vector Machine mampu menghasilkan model klasifikasi yang baik meskipun dilatih dengan himpunan data yang relatif sedikit hanya dengan pengaturan parameter yang sederhana. K-Nearest Neighbor merupakan algoritma yang menentukan nilai jarak pada pengujian data testing dengan data training berdasarkan nilai yang terkecil dari nilai ketetanggaan yang terdekat. K-Nearest Neighbor digunakan karena memiliki akurasi yang tinggi dengan kesalahan yang kecil.

II. LANDASAN TEORI

A. Data Mining

Data yaitu kumpulan fakta yang terekam atau sebuah entitas yang tidak memiliki arti dan selama ini terabaikan sedangkan *mining* yaitu proses penambangan. Sehingga *data mining* dapat diartikan proses penambangan data yang menghasilkan output (keluaran) yang berupa pengetahuan (Dicky Nofriansyah, 2015). *Data Mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data (Retno Tri Vulansari, 2017). *Data Mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya. (Pramudiono, 2006). *Data Mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. (Larose, 2005). *Data Mining* ditujukan untuk mengekstrak (mengambil intisari) pengetahuan dari sekumpulan data sehingga didapatkan struktur yang dapat dimengerti manusia serta meliputi basisdata dan management data, prapemrosesan data, pertimbangan model dan inferensi, ukuran ketertarikan, pertimbangan kompleksitas, pasca pemrosesan terhadap struktur yang ditemukan, visualisasi, dan *online updating* (ACM, 2006).

B. Naïve Bayes

Metode ini menggunakan Teorema Bayes, yang ditemukan oleh Thomas Bayes di abad ke-18. Teori Bayes adalah pendekatan statistika yang fundamental dalam *data mining*. Pendekatan ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas. Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema bayes (aturan bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat (Suyanto, 2019). Formula umum *Bayes* adalah sebagai berikut:

$$\frac{P(H|E)=P(E|H) \times P(H)}{P(E)}$$

Penjelasan dari formula tersebut adalah sebagai berikut:

P(H|E): Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis terjadi jika diberikan bukti (*evidence*) E terjadi.

P(E|H): Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H.

P(H): Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.

P(E): Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain.

C. Vector Support Machine

Support Vector Machine (SVM) dikembangkan oleh Boser, Guyon, Vapnik, dan pertama kali dipresentasikan pada tahun 1992 di Annual Workshop on Computational Learning Theory. Konsep dasar SVM sebenarnya merupakan kombinasi harmonis dari teori-teori komputasi yang telah ada puluhan tahun sebelumnya, seperti margin hyperplane (Duda & Hart tahun 1973, Cover tahun 1965, Vapnik 1964), kernel diperkenalkan oleh Aronsjan tahun 1950, dan demikian juga dengan konsep-konsep pendukung yang lain. Akan tetapi hingga tahun 1992, belum pernah ada upaya merangkaikan komponen-komponen tersebut. Berbeda dengan strategi Artificial Neural Network yang berusaha mencari hyperplane pemisah antar-class, SVM berusaha menemukan hyperplane yang terbaik pada input space.

Prinsip dasar SVM adalah linear classifier, dan selanjutnya dikembangkan agar dapat bekerja pada problem non-linear. dengan memasukkan konsep kernel trick pada ruang kerja berdimensi tinggi. Perkembangan ini memberikan rangsangan minat penelitian di bidang *pattern recognition* untuk investigasi potensi kemampuan SVM secara teoritis maupun dari segi aplikasi. Dewasa ini SVM telah berhasil diaplikasikan dalam problema dunia nyata (real-world problems), dan secara umum memberikan solusi yang lebih baik dibandingkan metode konvensional seperti misalnya artificial neural network (Nugroho, 2018).

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah class pada input space. Untuk n-dimensional space, input data x_i ($i=1. . .k$), dimana milik kelas 1 atau kelas 2 dan label yang terkait menjadi -1 untuk kelas 1 dan +1 untuk kelas 2. Gambar 1. memperlihatkan beberapa pattern yang merupakan anggota dari dua buah class: positif (dinotasikan dengan +1) dan negatif (dinotasikan dengan -1). Pattern yang tergabung pada class negatif disimbolkan dengan kotak, sedangkan pattern pada class positif, disimbolkan dengan lingkaran. Jika data input dapat dipisahkan secara linear, pemisahan hyperplane dapat diberikan dalam: Proses pembelajaran dalam problem klasifikasi diterjemahkan sebagai upaya menemukan garis (hyperplane) yang memisahkan antara kedua kelompok tersebut. SVM Berusaha Menemukan Hyperplane Terbaik yang Memisahkan Kedua Class Negatif dan Positif.

D. K-Nearest Neighbor

K-Nearest Neighborhood (k-NN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari label class pada k-NN (Suyanto, 2018).

Tujuan dari algoritma k-NN adalah mengklasifikasikan objek baru berdasarkan atribut dan training data. Algoritma k-NN bekerja berdasarkan jarak terpendek dari query instance ke training data untuk menentukan k-NN-nya. Salah satu cara untuk menghitung jarak dekat atau jauhnya tetangga menggunakan metode euclidean distance. Euclidean Distance sering digunakan untuk menghitung jarak. Euclidean Distance berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua objek, di bawah ini merupakan rumus Euclidean Distance:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

Keterangan:

d (xi, xj): Jarak Euclidean (Euclidean Distance).

(xi): record ke-i

(xj): record ke-j

(ar) : data ke-r

i, j:1,2,3,...n

Jika hasil nilai dari rumus di atas besar maka akan semakin jauh tingkat keserupaan antara kedua objek dan sebaliknya jika hasil nilainya semakin kecil maka akan semakin dekat tingkat keserupaan antar objek tersebut. Objek yang dimaksud adalah training data dan testing data. Langkah-langkah untuk menghitung algoritma k-NN:

- a. Menentukan nilai k.
- b. Menghitung kuadrat jarak euclid (query instance) masing-masing objek terhadap training data yang diberikan.
- c. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil.
- d. Mengumpulkan label class Y (klasifikasi Nearest Neighborhood).
- e. Dengan menggunakan kategori Nearest Neighborhood yang paling mayoritas maka dapat diprediksikan nilai query instance yang telah dihitung.

III. METODE PENELITIAN

A. Kategori Penelitian

Penelitian secara umum dapat diartikan sebagai proses pengumpulan dan analisis data yang dilakukan secara sistematis dan logis untuk mencapai tujuan tertentu dan memperkaya pengetahuan itu sendiri oleh penemuan fakta dan wawasan yang tidak biasa. Kategori penelitian ini adalah penelitian deskriptif dengan pendekatan kuantitatif dan analisis kausalitas. Populasi adalah wilayah generalisasi yang terdiri atas objek atau subjek yang mempunyai kualitas atau karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari kemudian ditarik kesimpulannya (Sudaryono, 2019). Populasi data dalam penelitian ini adalah 1233 data harga minyak kelapa sawit pada periode bulan agustus 2014 – bulan agustus 2019. Data diperoleh dari situs di <http://www.investing.com/>. Sample merupakan suatu bagian dari populasi, mencakup sejumlah anggota yang dipilih dari populasi (Sudaryono, 2019). Pengambilan

sample pada penelitian ini yaitu harga minyak kelapa sawit pada bulan agustus 2019.

B. Kerangka Pemikiran

Kerangka pemikiran dalam penelitian ini dibuat dengan tujuan agar penelitian dapat dilakukan secara bertahap serta konsisten dan merupakan garis besar dari langkah-langkah penelitian yang dilakukan dalam penelitian ini.



Gambar 1 Kerangka Pemikiran

C. Tahapan Penelitian

Pengumpulan data adalah cara atau teknik yang dapat digunakan oleh peneliti untuk mengumpulkan data. Atribut atau variabel adalah sifat atau nilai dari orang, objek, atau kegiatan yang mempunyai variasi tertentu yang ditetapkan oleh peneliti untuk dipelajari kemudian ditarik kesimpulannya (Sudaryono, 2015). Teknik pengumpulan data yang digunakan dalam penelitian ini adalah dokumentasi dan studi pustaka dengan uraian sebagai berikut:

1. Teknik pengumpulan data menggunakan teknik studi pustaka melalui pengumpulan informasi pada literature-literatur yang relevan untuk mendukung pembahasan penelitian. Studi pustaka dapat melalui buku teks/*e-books*, jurnal, ilmiah asing/*e-journal*, karya tulis ilmiah, thesis, media berita online, Yahoo Finance dan lainnya. Teknik pengumpulan data menggunakan teknik dokumentasi berdasarkan <http://www.investing.com/> pada periode bulan agustus 2014 – bulan agustus 2019 dengan atribut atau variabel sebagai berikut:

Tabel 1 Atribut atau Variabel yang Digunakan

No	Nama Atribut/variabel	Keterangan
1.	Minyak Kelapa Sawit	Merupakan data harga minyak kelapa sawit pertahun.
2.	Minyak Kedelai	Merupakan data harga minyak kedelai pertahun.
3.	Minyak Canola	Merupakan data harga minyak canola pertahun.

No	Nama Atribut/variabel	Keterangan
4.	Harga Minyak Mentah WTI	Merupakan data harga minyak mentah dunia pertahun.
5.	USD_IDR	Merupakan nilai tukar riil antara kurs dolar terhadap rupiah.
6.	USD_MYR	Merupakan nilai tukar riil antara kurs dolar terhadap ringgit.
7.	USD_INR	Merupakan nilai tukar riil antara kurs dolar terhadap rupee.
8.	USD_CAD	Merupakan nilai tukar riil antara kurs dolar terhadap dolar canada.

2. Pengolahan data pada penelitian ini menggunakan teknik klasifikasi pada data mining untuk memprediksi harga minyak kelapa sawit dalam investasi dengan membandingkan algoritma Naïve Bayes, Support Vector Machine, dan K-Nearest Neighbor. Data tersebut akan dibagi menjadi 90% data *training* dan 10% data *testing*. Memisahkan data menjadi *training* dan *testing set* dimaksudkan agar model yang diperoleh nantinya memiliki kemampuan generalisasi yang baik dalam melakukan klasifikasi data. Data *training* atau *training set* adalah bagian dataset yang dilatih untuk membuat prediksi atau menjalankan fungsi dari sebuah algoritma sesuai dengan tujuannya masing-masing. Data *testing* atau *test set* adalah bagian dataset yang digunakan untuk melihat keakuratan atau performa.
3. Data Selection merupakan proses pemilihan data dari sekumpulan data operasional yang ada sebelum masuk ke tahap mining data maupun informasi. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.
4. Pada tahap *data cleaning* merupakan proses pembersihan dari data yang akan dipakai untuk penghapusan data dengan membuang *missing value*, duplikasi data, dan memeriksa inkonsistensi data dan memperbaiki kesalahan pada data. Proses pembersihan data dilakukan secara manual dengan bantuan *software spreadsheet*.
5. Tahap *Data Transformation* merupakan proses mengubah format data awal menjadi sebuah format data standar untuk proses pembacaan data pada program maupun *tool* yang digunakan.

Tabel 2 Transformasi Data Harga Minyak Kelapa Sawit

Minyak Sawit	Minyak Kedelai	Minyak Canola	Minyak Mentah WTI	USD_IDR	USD_MYR	USD_INR	USD_CAD
514,2	28,64	448	55,1	1418	4,21	71,45	1,33

Minyak Sawit	Minyak Kedelai	Minyak Canola	Minyak Mentah WTI	USD_IDR	USD_MYR	USD_INR	USD_CAD
5				5			
511	28,37	447,5	56,71	14237,5	4,22	71,69	1,33
522,25	28,11	447,9	55,78	14255	4,21	71,83	1,33
519,25	28,01	446,4	54,93	14255	4,2	71,54	1,33
526,5	28,33	448,4	53,64	14240	4,2	72,03	1,33
527,5	28,34	449,9	54,17	14212,5	4,19	71,51	1,33
525,75	28,54	453,1	55,35	14235	4,19	71,93	1,33
517,5	28,74	453,1	55,68	14240	4,18	71,44	1,33
509	28,43	451,4	56,34	14260	4,18	71,55	1,33
504	28,68	450,2	56,21	14235	4,18	71,45	1,33
512,4	29,13	452,4	54,87	14235	4,18	71,16	1,33
512,5	29,07	451,2	54,47	14267,5	4,19	71,81	1,33
519,25	29,57	450,2	55,23	14245	4,19	71,28	1,33
517,5	29,61	448,9	57,1	14317,5	4,2	71,42	1,32
511,75	29,51	452,4	54,5	14190	4,18	70,92	1,32
495,75	28,96	454,1	52,54	14208	4,19	70,42	1,32
487,25	27,96	452,2	51,09	14222,5	4,19	71,15	1,33
486,25	27,48	449,4	53,63	14265	4,19	71,04	1,33
485,25	27,76	448	54,69	14255	4,18	70,82	1,32
484,25	28,19	444,8	55,66	14180	4,16	69,7	1,32
486,25	27,68	442	53,95	14116	4,15	69,02	1,32

D. Pemodelan

Pemodelan pada penelitian ini dilakukan dengan *data mining* teknik klasifikasi membandingkan algoritma Naïve Bayes, Support Vector Machine, dan K-Nearest Neighbor.

1. Pemodelan Perhitungan Naïve Bayes :

P(H):

$P(\text{prediksi} = \text{"naik"}) = 13/21 = 0.619$, $P(\text{prediksi} = \text{"turun"}) = 8/21 = 0.381$

Training: Hitung $P(E|H)$ untuk setiap kelas

$P(\text{Minyak Sawit} = \text{"<=515"} | \text{prediksi} = \text{"naik"}) = 7/13 = 0.538$

$P(\text{Minyak Sawit} = \text{"<=515"} | \text{prediksi} = \text{"turun"}) = 6/8 = 0.75$

$P(\text{Minyak Sawit} = \text{">=515"} | \text{prediksi} = \text{"naik"}) = 6/13 = 0.461$

$P(\text{Minyak Sawit} = \text{">=515"} | \text{prediksi} = \text{"turun"}) = 2/8 = 0.25$

$P(\text{Minyak Kedelai} = \text{"<=28"} | \text{prediksi} = \text{"naik"}) = 3/13 = 0.231$

$P(\text{Minyak Kedelai} = \text{"<=28"} | \text{prediksi} = \text{"turun"}) = 1/8 = 0.125$

$P(\text{Minyak Kedelai} = \text{">=28"} | \text{prediksi} = \text{"naik"}) = 10/13 = 0.769$

$P(\text{Minyak Kedelai} = \text{">=28"} | \text{prediksi} = \text{"turun"}) = 7/8 = 0.875$

$P(\text{Minyak Canola} = \text{"<=450"} | \text{prediksi} = \text{"naik"}) = 6/13 = 0.461$

$P(\text{Minyak Canola} = \text{"<=450"} | \text{prediksi} = \text{"turun"}) = 5/8 = 0.625$

$P(\text{Minyak Canola} = \text{">=450"} | \text{prediksi} = \text{"naik"}) = 7/13 = 0.538$

$P(\text{Minyak Canola} = \text{">=450"} | \text{prediksi} = \text{"turun"}) = 3/8 = 0.375$

$P(\text{Minyak Mentah WTI} = \text{"<=55"} | \text{prediksi} = \text{"naik"}) = 6/13 = 0.461$

$P(\text{Minyak Mentah WTI} = \text{"<=55"} | \text{prediksi} = \text{"turun"}) = 5/8 = 0.625$

$P(\text{Minyak Mentah WTI} = \text{">=55"} | \text{prediksi} = \text{"naik"}) = 7/13 = 0.538$

$P(\text{Minyak Mentah WTI} = \text{">=55"} | \text{prediksi} = \text{"turun"}) = 3/8 = 0.375$

$P(\text{USD_IDR} = \text{"<=14220"} | \text{prediksi} = \text{"naik"}) = 4/13 = 0.308$

$P(\text{USD_IDR} = \text{"<=14220"} | \text{prediksi} = \text{"turun"}) = 2/8 = 0.25$

$P(\text{USD_IDR} = \text{">=14220"} | \text{prediksi} = \text{"naik"}) = 9/13 = 0.692$

$P(\text{USD_IDR} = \text{">=14220"} | \text{prediksi} = \text{"turun"}) = 6/8 = 0.75$

$P(\text{USD_MYR} = \text{"<=4.19"} | \text{prediksi} = \text{"naik"}) = 10/13 = 0.769$

$P(\text{USD_MYR} = \text{"<=4.19"} | \text{prediksi} = \text{"turun"}) = 5/8 = 0.625$

$P(\text{USD_MYR} = \text{">=4.19"} | \text{prediksi} = \text{"naik"}) = 3/13 = 0.231$

$P(\text{USD_MYR} = \text{">=4.19"} | \text{prediksi} = \text{"turun"}) = 3/8 = 0.375$

$P(\text{USD_INR} = \text{"<=71.6"} | \text{prediksi} = \text{"naik"}) = 11/13 = 0.846$

$P(\text{USD_INR} = \text{"<=71.6"} | \text{prediksi} = \text{"turun"}) = 5/8 = 0.625$

$P(\text{USD_INR} = \text{">=71.6"} | \text{prediksi} = \text{"naik"}) = 2/13 = 0.154$

$P(\text{USD_INR} = \text{">=71.6"} | \text{prediksi} = \text{"turun"}) = 3/8 = 0.375$

$P(\text{USD_CAD} = \text{"<=1.31"} | \text{prediksi} = \text{"naik"}) = 0/13 = 0$

$P(\text{USD_CAD} = \text{"<=1.31"} | \text{prediksi} = \text{"turun"}) = 0/8 = 0$

$P(\text{USD_CAD} = \text{">=1.31"} | \text{prediksi} = \text{"naik"}) = 13/13 = 1$

$P(\text{USD_CAD} = \text{">=1.31"} | \text{prediksi} = \text{"turun"}) = 8/8 = 1$
 Klasifikasi: $E = (\text{Minyak Sawit} = \text{"<=515"}, P(\text{Minyak Kedelai} = \text{">=28"}, \text{Minyak Canola} = \text{"<=450"}, \text{Minyak Mentah WTI} = \text{">=55"}, \text{USD_IDR} = \text{"<=14220"}, \text{USD_MYR} = \text{">=4.19"}, \text{USD_INR} = \text{"<=71.6"}, \text{USD_CAD} = \text{">=1.31"})$

$P(E|H)$:

$P(E|\text{prediksi} = \text{"naik"}) = 0.538 \times 0.769 \times 0.461 \times 0.538 \times 0.308 \times 0.231 \times 0.846 \times 1 = 0.006$, $P(E|\text{prediksi} = \text{"turun"}) = 0.75 \times 0.875 \times 0.625 \times 0.375 \times 0.25 \times 0.375 \times 0.625 \times 1 = 0.009$

$P(E|H)*P(H)$:

$P(E|\text{prediksi} = \text{"naik"}) * P(\text{prediksi} = \text{"naik"}) = 0.006 * 0.619 = 0.004$

$P(E|\text{prediksi} = \text{"turun"}) * P(\text{prediksi} = \text{"turun"}) = 0.009 * 0.381 = 0.003$

Kesimpulan: $0.004 > 0.003$ maka prediksi = "naik".

2. Pemodelan Perhitungan Support Vector Machine:

Berdasarkan Tabel dan grafik hyperplane diperoleh persamaan *constraint* (pembatas) sebagai berikut:

$$514,25W_1 + 28,64W_2 + b \geq 1$$

$$-511W_1 - 28,37W_2 - b \geq 1$$

$$522,25W_1 + 28,11W_2 + b \geq 1$$

$$-519,25W_1 - 28,01W_2 - b \geq 1$$

$$526,5W_1 + 28,33W_2 + b \geq 1$$

Eliminasi persamaan diperoleh:

$$[1] \quad 514,25W_1 + 28,64W_2 + b = 1$$

$$[2] \quad -511W_1 - 28,37W_2 - b = 1 \quad +$$

$$3,25W_1 + 0,27W_2 = 2$$

$$3,25W_1 = 2 - 0,27W_2$$

$$W_1 = 1,73 - 0,27W_2$$

$$[3] \quad -3,25W_1 - 0,27W_2 - b = 1$$

$$[4] \quad 522,25W_1 + 28,11W_2 + b = 1 \quad +$$

$$519W_1 + 27,84W_2 = 2$$

$$897,87 - 1686,75W_2 + 27,84W_2 = 2$$

$$-761,04W_2 = -4$$

$$W_2 = 0,005, \text{ maka } W_1 = 1,73 - 0,01 = 1,72$$

Substitusi nilai W_1 dan W_2 kedalam persamaan:

$$[5] \quad -519,25W_1 - 28,01W_2 - b = 1$$

$$-519,25 * (1,72) - 28,01 * (0,005) - b = 1$$

$$-893,11 - (0,14005) - b = 1$$

$$b = 1 + (-893,25)$$

$$b = -892,25$$

berdasarkan nilai W_1 , W_2 , b yang diperoleh, maka persamaan fungsi pemisahannya yaitu:

$$f(x) = 1,72x_1 + 0,005x_2 - 892,25$$

$$\text{Jika } x_1 = 0, x_2 = 895,25/0,005$$

$$\text{Jika } x_2 = 0, x_1 = -895,25/1,72$$

klasifikasi untuk data selanjutnya dapat dilakukan dengan menggunakan fungsi pada persamaan, dengan fungsi tersebut semua nilai $f(x) < 0$ masuk kekelas -1 (negatif) dan nilai $f(x) > 0$ masuk kekelas +1 (positif).

3. Pemodelan Perhitungan K-Nearest Neighbor:

$$\text{Jarak} = \sqrt{((514.25-514)^2 + (28.64-29)^2 + (448-447)^2 + (55.1-56)^2 + (14185-14200)^2 + (4.21-4.2)^2 + (71.45-71.50)^2 + (1.33-1.32)^2)}$$

$$\text{Jarak} = \sqrt{((0.25)^2 + (-0.36)^2 + (1)^2 + (-0.9)^2 + (-15)^2 + (0.01)^2 + (-0.05)^2 + (1)^2)}$$

$$\text{Jarak} = \sqrt{(227.0048)}$$

$$\text{Jarak} = 15.066678$$

E. Evaluasi dan Validasi Model

Validasi model k-fold cross validation dilakukan dengan membagi data menjadi bilangan bulat sejumlah k. Nilai k yang digunakan pada penelitian ini adalah 10. Sehingga dilakukan proses pelatihan dan pengujian sebanyak 10 kali. Dari 1233 data (5 tahun) yang digunakan pada penelitian ini, jika dibagi 10 bagian maka setiap bagiannya terdiri dari data per 6 bulan. Pembagian data untuk dijadikan data training dan data testing terus dilakukan hingga sepuluh kali.

Tabel 3 Validasi model k-fold cross validation

1233 Data (5 Tahun)										
No	A (Data 6 Bulan)	B (Data 6 Bulan)	C (Data 6 Bulan)	D (Data 6 Bulan)	E (Data 6 Bulan)	F (Data 6 Bulan)	G (Data 6 Bulan)	H (Data 6 Bulan)	I (Data 6 Bulan)	J (Data 6 Bulan)
1	A (Data Uji)	BCDEFGHIJ (Data Latih)								
2	A (Data Latih)	B (Data Uji)	CDEFGHIJ (Data Latih)							
3	AB (Data Latih)		C (Data Uji)	DEFGHIJ (Data Latih)						
4	ABC (Data Latih)			D (Data Uji)	EFGHIJ (Data Latih)					
5	ABCD (Data Latih)				E (Data Uji)	FGHIJ (Data Latih)				
6	ABCDE (Data Latih)					F (Data Uji)	GHIJ (Data Latih)			
7	ABCDEF (Data Latih)						G (Data Uji)	HIJ (Data Latih)		
8	ABCDEFG (Data Latih)							H (Data Uji)	IJ (Data Latih)	
9	ABCDEFGH (Data Latih)								I (Data Uji)	J (Data Latih)
10	ABCDEFGHI (Data Latih)									J (Data Uji)

Validasi metode dan algoritma Naïve Bayes, Support Vector Machine dan K-Nearest Neighbor dilakukan dengan mengukur evaluasi hasil *accuracy*, *precision* dan *recall* dapat dihitung dengan menggunakan *Confusion Matrix* pada data sample bulan agustus 2019 sebagai berikut:

1. Naïve Bayes

Tabel 4 Confusion Matrix Perhitungan Naïve Bayes

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP 10	FP 6
	FALSE	FN 3	TN 2

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

$$= \frac{10 + 2}{10 + 2 + 6 + 3} * 100\%$$

$$= \frac{12}{21} * 100\%$$

$$= 0,5714 * 100\%$$

$$= 57,14\%$$

$$\text{Precision} = \frac{TP}{TP + FP} * 100\%$$

$$= \frac{10}{10 + 6} * 100\%$$

$$= \frac{10}{16} * 100\%$$

$$= 0,625 * 100\%$$

$$= 62,5\%$$

$$\text{Recall} = \frac{TP}{TP + FN} * 100\%$$

$$= \frac{10}{10 + 3} * 100\%$$

$$= \frac{10}{13} * 100\%$$

$$= 0,6923 * 100\%$$

$$= 76,92\%$$

2. Support Vector Machine

Tabel 5 Confusion Matrix Perhitungan Support Vector Machine

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP 12	FP 6
	FALSE	FN 1	TN 2

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

$$= \frac{12 + 2}{12 + 2 + 6 + 1} * 100\%$$

$$= \frac{14}{21} * 100\%$$

$$= 0,6667 * 100\%$$

$$= 66,67\%$$

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} * 100\% \\
 &= \frac{12}{12 + 6} * 100\% \\
 &= \frac{12}{18} * 100\% \\
 &= 0,6667 * 100\% \\
 &= 66,67\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} * 100\% \\
 &= \frac{12}{12 + 1} * 100\% \\
 &= \frac{12}{13} * 100\% \\
 &= 92,31 * 100\% \\
 &= 92,31\%
 \end{aligned}$$

3. K-Nearest Neighbor

**Tabel 6 Confusion Matrix
Perhitungan K-Nearest Neighbor**

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP 6	FP 7
	FALSE	FN 7	TN 1

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} * 100\% \\
 &= \frac{6 + 1}{6 + 1 + 7 + 7} * 100\% \\
 &= \frac{7}{21} * 100\% \\
 &= 0,3333 * 100\% \\
 &= 33,33\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} * 100\% \\
 &= \frac{6}{6 + 7} * 100\% \\
 &= \frac{6}{13} * 100\% \\
 &= 0,4615 * 100\% \\
 &= 46,15\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} * 100\% \\
 &= \frac{6}{6 + 7} * 100\% \\
 &= \frac{6}{13} * 100\% \\
 &= 0,4615 * 100\% \\
 &= 46,15\%
 \end{aligned}$$

Tabel 7 Hasil pengujian data sampel (Agustus 2019)

No	Algoritma	Accuracy	Precision	Recall
1	Naïve Bayes	52,38%	60%	69,23%
2	SVM	66,67%	66,67%	92,31%
3	K-NN	33,33%	46,15%	46,15%

IV. HASIL DAN PEMBAHASAN

Objek dari penelitian ini adalah pengujian terhadap perbandingan algoritma Naïve Bayes, Support Vector Machine, dan K-Nearest Neighbor untuk mengetahui algoritma yang mana yang akan mendapatkan hasil nilai *accuracy*, *precision*, dan *recall* serta prediksi yang dapat digunakan dalam pengambilan keputusan ketika berinvestasi minyak kelapa sawit. Sumber data sebagai objek pada penelitian ini adalah data historis harga minyak kelapa sawit dalam kurun waktu 5 tahun yang diambil dari situs <http://www.investing.com/> pada periode bulan agustus 2014 – bulan agustus 2019. Data yang digunakan dalam penelitian ini terdiri dari atribut atau variabel seperti: harga minyak kelapa sawit, harga minyak kedelai, harga minyak canola, harga minyak mentah WTI, USD_IDR, USD_MYR, USD_INR, dan USD_CAD.

A. Hasil Uji Penelitian

Menggunakan fungsi klasifikasi data yang jumlahnya besar dikelompokkan menjadi data yang lebih kecil-kecil. Maka setelah diklasifikasi, data tersebut dapat diolah dengan menggunakan program rapidminer versi 9.0 dengan tujuan untuk melihat *accuracy*, *precision*, dan *recall* dari setiap algoritma yaitu Naïve Bayes, KNN atau SVM mana yang memiliki akurasi paling baik dan mendekati sempurna. Hasil pengolahan data set sebanyak 1233 data selama 5 tahun dibagi 10 bagian maka setiap bagiannya terdiri dari data *testing* selama 6 bulan.

Tabel 8 Summary Result

No	Algoritma	Accuracy	Precision	Recall
1a	Naïve Bayes	65,79%	67,35%	58,93%
	SVM	66,67%	71,43%	53,57%
	K-NN	63,16%	62,96%	60,71%
1b	Naïve Bayes	56,14%	54,84%	60,71%
	SVM	52,63%	51,47%	62,50%
	K-NN	57,89%	57,41%	55,36%
1c	Naïve Bayes	58,77%	63,64%	37,50%
	SVM	69,30%	67,80%	71,43%
	K-NN	68,42%	66,67%	71,43%
1d	Naïve Bayes	64,91%	71,05%	48,21%
	SVM	82,46%	86%	76,79%
	K-NN	68,42%	66,67%	71,43%
2a	Naïve Bayes	65,85%	68,49%	72,46%
	SVM	63,41%	65,38%	73,91%
	K-NN	60,98%	65,22%	65,22%
2b	Naïve Bayes	60,16%	65,62%	60,87%
	SVM	60,16%	66,67%	57,97%
	K-NN	60,16%	63,51%	68,12%
2c	Naïve Bayes	67,48%	75,44%	62,32%
	SVM	73,17%	74,32%	79,71%

No	Algoritma	Accuracy	Precision	Recall
	K-NN	58,54%	61,84%	68,12%
2d	Naïve Bayes	67,48%	74,58%	63,77%
	SVM	77,24%	78,87%	81,16%
	K-NN	62,60%	66,67%	66,67%
3a	Naïve Bayes	65,29%	67,19%	67,19%
	SVM	68,60%	67,11%	79,69%
	K-NN	66,94%	66,22%	76,56%
3b	Naïve Bayes	52,89%	54,67%	64,06%
	SVM	50,41%	52,63%	62,50%
	K-NN	60,33%	60%	75%
3c	Naïve Bayes	66,94%	78,57%	51,56%
	SVM	66,12%	69,49%	64,06%
	K-NN	60,33%	64,29%	56,25%
3d	Naïve Bayes	66,94%	71,43%	62,50%
	SVM	81,82%	79,17%	89,06%
	K-NN	60,33%	63,79%	57,81%
4a	Naïve Bayes	54,47%	56,90%	51,56%
	SVM	60,16%	59,49%	73,44%
	K-NN	53,66%	55,56%	54,69%
4b	Naïve Bayes	50,41%	52,31%	53,12%
	SVM	50,41%	52,11%	57,81%
	K-NN	45,53%	47,62%	46,88%
4c	Naïve Bayes	65,85%	72,92%	54,69%
	SVM	68,29%	75,51%	57,81%
	K-NN	64,23%	66,13%	64,06%
4d	Naïve Bayes	65,85%	73,91%	53,12%
	SVM	78,86%	85,19%	71,88%
	K-NN	68,29%	69,23%	70,31%
5a	Naïve Bayes	58,20%	58,33%	66,67%
	SVM	61,48%	61,11%	69,84%
	K-NN	54,92%	56,25%	57,14%
5b	Naïve Bayes	52,46%	52,81%	74,60%
	SVM	50,82%	52,31%	53,97%
	K-NN	54,10%	54,79%	63,49%
5c	Naïve Bayes	66,39%	73,91%	53,97%
	SVM	70,49%	72,13%	69,84%
	K-NN	65,57%	69,81%	58,73%
5d	Naïve Bayes	70,49%	76,47%	61,90%
	SVM	72,13%	73,77%	71,43%
	K-NN	64,75%	68,52%	58,73%
6a	Naïve Bayes	57,72%	62,50%	33,33%
	SVM	52,85%	52,17%	40%
	K-NN	59,35%	58,62%	56,67%
6b	Naïve Bayes	61,79%	63,83%	50%
	SVM	53,66%	52,31%	56,67%
	K-NN	62,60%	64,58%	51,67%
6c	Naïve Bayes	68,29%	65,22%	75%
	SVM	72,36%	72,41%	70%
	K-NN	60,16%	58,73%	61,67%
6d	Naïve Bayes	71,54%	69,23%	75%
	SVM	75,61%	76,79%	71,67%
	K-NN	60,16%	59,02%	60%
7a	Naïve Bayes	64,23%	62,50%	72,58%
	SVM	62,60%	63,79%	59,68%
	K-NN	54,47%	54,84%	54,84%
7b	Naïve Bayes	52,85%	51,96%	85,48%

No	Algoritma	Accuracy	Precision	Recall
	SVM	52,85%	52%	83,87%
	K-NN	57,72%	56,76%	67,74%
7c	Naïve Bayes	64,23%	66,07%	59,68%
	SVM	68,29%	66,20%	75,81%
	K-NN	62,60%	64,29%	58,06%
7d	Naïve Bayes	72,36%	71,88%	74,19%
	SVM	73,98%	74,19%	74,19%
	K-NN	60,98%	62,07%	58,06%
8a	Naïve Bayes	63,03%	60%	72,41%
	SVM	66,39%	68%	58,62%
	K-NN	68,91%	67,80%	68,97%
8b	Naïve Bayes	50,42%	49,44%	75,86%
	SVM	54,62%	52,63%	68,97%
	K-NN	52,10%	50,88%	50%
8c	Naïve Bayes	63,87%	62,30%	65,52%
	SVM	66,39%	64,52%	68,97%
	K-NN	63,87%	63,64%	60,34%
8d	Naïve Bayes	68,07%	67,24%	67,24%
	SVM	78,15%	77,59%	77,59%
	K-NN	63,87%	63,64%	60,34%
9a	Naïve Bayes	63,93%	63,38%	71,43%
	SVM	63,11%	63,64%	66,67%
	K-NN	60,66%	62,30%	60,32%
9b	Naïve Bayes	56,56%	59,26%	50,79%
	SVM	53,28%	54,84%	53,97%
	K-NN	63,11%	66,07%	58,73%
9c	Naïve Bayes	72,95%	69,74%	84,13%
	SVM	70,49%	69,57%	76,19%
	K-NN	64,75%	64,71%	69,84%
9d	Naïve Bayes	71,31%	73,33%	69,84%
	SVM	76,23%	75%	80,95%
	K-NN	63,11%	63,64%	66,67%
10a	Naïve Bayes	54,92%	52,46%	55,17%
	SVM	55,74%	53,45%	53,45%
	K-NN	55,74%	53,12%	58,62%
10b	Naïve Bayes	54,10%	51,85%	48,28%
	SVM	54,92%	52,54%	53,45%
	K-NN	52,46%	50%	46,55%
10c	Naïve Bayes	65,57%	65,38%	58,62%
	SVM	68,85%	67,24%	67,24%
	K-NN	68,85%	67,86%	65,52%
10d	Naïve Bayes	64,75%	63,16%	62,07%
	SVM	76,23%	75,44%	74,14%
	K-NN	68,03%	66,10%	67,24%

B. Analisis Hasil Penelitian

Analisis hasil dari penelitian ini merupakan rangkuman hasil percobaan dari setiap algoritma dan salah satu dari algoritma tersebut akan dipilih oleh mana yang paling baik. Dari tabel 8 dengan melakukan beberapa kali pengujian algoritma berbeda-beda yaitu Naïve Bayes, Support Vector Machine dan K-Nearest Neighbor didapatkan hasil *accuracy*, *precision*, dan *recall* yang berbeda-beda.

C. Atribut yang Berpengaruh Signifikan dan Atribut yang Tidak Berpengaruh Signifikan

Berdasarkan hasil pengujian yang sudah dilakukan dalam penelitian ini yaitu atribut harga minyak nabati lain (minyak kedelai dan minyak canola) dan atribut nilai tukar riil antara kurs dolar terhadap mata uang negara produsen (rupiah, ringgit, dan canada) atau mata uang negara konsumen (rupee) berpengaruh signifikan dalam penelitian ini terbukti memiliki *accuracy*, *precision*, dan *recall* yang paling tinggi apabila menggunakan atribut tersebut pada pengujian tabel 4.1d, tabel 4.3a, tabel 4.3c, tabel 4.3d, tabel 4.5c, tabel 4.8a, dan tabel 4.9c. Sedangkan atribut harga minyak mentah dunia tidak berpengaruh signifikan dalam penelitian ini terbukti memiliki *accuracy*, *precision*, dan *recall* yang paling rendah apabila menggunakan atribut tersebut pada pengujian tabel 4.1b, tabel 4.3b, tabel 4.4b, dan 4.8b.

D. Faktor yang Mempengaruhi Nilai Selisih Algoritma

Dari hasil pengujian beberapa data dan atribut, selisih *accuracy* antara algoritma Support Vector Machine dengan Naïve Bayes adalah 9,51% sedangkan Selisih *accuracy* antara algoritma Support Vector Machine dengan K-Nearest Neighbor adalah 13,55%. Begitu pula selisih *precision* antara algoritma Support Vector Machine dengan Naïve Bayes adalah 7,43% sedangkan Selisih *precision* antara algoritma Support Vector Machine dengan K-Nearest Neighbor adalah 16,19%. Sedangkan selisih *recall* antara algoritma Support Vector Machine dengan Naïve Bayes adalah 3,58% dan selisih *recall* antara algoritma Support Vector Machine dengan K-Nearest Neighbor adalah 12,5%.

Hal ini dikarenakan kenaikan atau penurunan perbandingan angka pervariabel sangat berbeda antara variabel yang satu dengan yang lainnya. Dataset yang diolah dalam penelitian ini berupa angka yang perbandingan nominalnya sangat berbeda antara variabel yang satu dengan variabel yang lain contohnya atribut dengan nominal satuan pada atribut USD_MYR dan USD_CAD, atribut dengan nominal puluhan pada atribut minyak kedelai, minyak mentah WTI dan USD_INR, atribut dengan nominal ratusan pada atribut minyak kelapa sawit dan minyak canola, sedangkan atribut dengan nominal ribuan pada atribut USD_IDR.

Dari semua hasil data testing dengan periode 6 bulan bahwa algoritma yang memiliki *accuracy*, *precision*, dan *recall* dengan jumlah paling tinggi adalah algoritma Support Vector Machine karena dalam penelitian ini hanya menggunakan dua kelas yaitu kelas naik dan kelas turun pada prediksi harga minyak kelapa sawit dalam investasi. Algoritma Support Vector Machine umumnya diformulasikan untuk menyelesaikan masalah klasifikasi dua kelas. Algoritma Support Vector Machine memiliki keunggulan dalam optimasi sistem pengenalan pola dengan kemampuan generalisasi yang baik. Algoritma Support Vector Machine mampu menghasilkan model klasifikasi yang baik meskipun dilatih dengan himpunan data yang relatif sedikit hanya dengan pengaturan parameter yang sederhana. Sedangkan algoritma K-Nearest Neighbor menghasilkan tingkat *accuracy*, *precision*, dan *recall* yang paling rendah karena pada penelitian ini jarak selisih angka

pada atribut yang satu dengan atribut yang lain terlalu besar contohnya atribut dengan nominal satuan pada atribut USD_MYR dan USD_CAD, atribut dengan nominal puluhan pada atribut minyak kedelai, minyak mentah WTI dan USD_INR, atribut dengan nominal ratusan pada atribut minyak kelapa sawit dan minyak canola, sedangkan atribut dengan nominal ribuan pada atribut USD_IDR. Sedangkan algoritma K-Nearest Neighbor dipengaruhi nilai atau jarak tetangga terdekat.

E. Data yang Terklasifikasi Salah

Data yang terklasifikasi salah dipengaruhi oleh nilai minimum (*min*), nilai maksimum (*max*) dan nilai rata-rata (*average*) pada setiap atribut. Apabila atribut pada data tersebut nilainya lebih kecil dari nilai rata-rata (*average*) maka prediksinya bernilai turun namun apabila atribut pada data tersebut nilainya lebih besar dari nilai rata-rata (*average*) maka prediksinya bernilai naik. Dataset yang diolah dalam penelitian ini berupa angka yang perbandingan nominalnya sangat berbeda antara atribut yang satu dengan atribut yang lain contohnya atribut dengan nominal satuan pada atribut USD_MYR dan USD_CAD, atribut dengan nominal puluhan pada atribut minyak kedelai, minyak mentah WTI dan USD_INR, atribut dengan nominal ratusan pada atribut minyak kelapa sawit dan minyak canola, sedangkan atribut dengan nominal ribuan pada atribut USD_IDR.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Dari semua hasil data testing dengan periode 6 bulan bahwa algoritma yang memiliki *accuracy*, *precision*, dan *recall* dengan jumlah paling tinggi adalah algoritma Support Vector Machine dengan pencapaian *accuracy* tertinggi adalah 82,46% pada tabel 4.1d dengan Atribut Harga Minyak Kelapa Sawit, Minyak Kedelai, Minyak Canola, Minyak Mentah WTI, USD_IDR, USD_MYR, USD_INR, dan USD_CAD (Agustus 2014 – Januari 2015) dan *precision* tertinggi adalah 86% pada tabel pada tabel 4.1d dengan Atribut Harga Minyak Kelapa Sawit, Minyak Kedelai, Minyak Canola, Minyak Mentah WTI, USD_IDR, USD_MYR, USD_INR, dan USD_CAD (Agustus 2014 – Januari 2015) serta *recall* tertinggi adalah 89,06% pada tabel 4.3d dengan Atribut Harga Minyak Kelapa Sawit, Minyak Kedelai, Minyak Canola, Minyak Mentah WTI, USD_IDR, USD_MYR, USD_INR, dan USD_CAD (Agustus 2015 – Januari 2016).

Atribut harga minyak nabati lain (minyak kedelai dan minyak canola) dan atribut nilai tukar riil antara kurs dolar terhadap mata uang negara produsen (rupiah, ringgit, dan canada) atau mata uang negara konsumen (rupee) berpengaruh signifikan dalam penelitian ini terbukti memiliki *accuracy*, *precision*, dan *recall* yang paling tinggi apabila menggunakan atribut tersebut pada pengujian tabel 4.1d, tabel 4.3a, tabel 4.3c, tabel 4.3d, tabel 4.5c, tabel 4.8a, dan tabel 4.9c. Sedangkan atribut harga minyak mentah dunia tidak berpengaruh signifikan dalam penelitian ini terbukti memiliki *accuracy*, *precision*, dan *recall* yang

paling rendah apabila menggunakan atribut tersebut pada pengujian tabel 4.1b, tabel 4.3b, tabel 4.4b, dan 4.8b.

Selisih *accuracy* antara algoritma Support Vector Machine dengan Naïve Bayes adalah 9,51% sedangkan Selisih *accuracy* antara algoritma Support Vector Machine dengan K-Nearest Neighbor adalah 13,55%. Begitu pula selisih *precision* antara algoritma Support Vector Machine dengan Naïve Bayes adalah 7,43% sedangkan Selisih *precision* antara algoritma Support Vector Machine dengan K-Nearest Neighbor adalah 16,19%. Sedangkan selisih *recall* antara algoritma Support Vector Machine dengan Naïve Bayes adalah 3,58% dan selisih *recall* antara algoritma Support Vector Machine dengan K-Nearest Neighbor adalah 12,5%. Hal ini dikarenakan kenaikan/penurunan perbandingan angka pervariabel sangat berbeda antara variabel yang satu dengan yang lainnya. Dataset yang diolah dalam penelitian ini berupa angka yang perbandingan nominalnya sangat berbeda antara variabel yang satu dengan variabel yang lain contohnya atribut dengan nominal satuan pada atribut USD_MYR dan USD_CAD, atribut dengan nominal puluhan pada atribut minyak kedelai, minyak mentah WTI dan USD_INR, atribut dengan nominal ratusan pada atribut minyak kelapa sawit dan minyak canola, sedangkan atribut dengan nominal ribuan pada atribut USD_IDR.

Data yang terklasifikasi salah dipengaruhi oleh nilai minimum (*min*), nilai maksimum (*max*) dan nilai rata-rata (*average*) pada setiap atribut. Apabila atribut pada data tersebut nilainya lebih kecil dari nilai rata-rata (*average*) maka prediksinya bernilai turun namun apabila atribut pada data tersebut nilainya lebih besar dari nilai rata-rata (*average*) maka prediksinya bernilai naik.

B. Saran

Berdasarkan hasil penelitian dan kesimpulan diatas, maka peneliti memberikan saran sebagai berikut:

1. Investor hendaknya lebih baik memperhatikan aspek harga-harga naik turunnya yang berkaitan dengan harga fluktuatif harga minyak kelapa sawit. Seperti dalam penelitian ini faktor yang mempengaruhi harga minyak kelapa sawit seperti harga minyak nabati lain (minyak kedelai dan minyak canola), harga minyak mentah dunia, maupun nilai tukar riil antara kurs dolar terhadap mata uang negara produsen (rupiah, ringgit, canada) atau mata uang negara konsumen (rupee) berpengaruh signifikan dalam melakukan prediksi harga minyak kelapa sawit.
2. Bagi peneliti selanjutnya yang ingin melakukan penelitian harga minyak kelapa sawit hendaknya menambahkan unsur-unsur penelitian lainnya sehingga dapat memberikan pengaruh yang lebih besar terhadap hasil penelitian. Unsur-unsur tersebut adalah:
 - a. Diharapkan penelitian berikutnya mampu meningkatkan trend *accuracy*, *precision*, dan *recall* prediksi yang lebih tinggi yaitu menggunakan model yang berbeda.
 - b. Menambahkan jumlah data set, sampel atau atribut penelitian, sehingga dapat meningkatkan hasil *accuracy*, *precision*, dan *recall*.

- c. Menambahkan variabel-variabel lain yang secara teori memiliki pengaruh terhadap prediksi harga minyak kelapa sawit.
- d. Menambahkan referensi-referensi baik itu dari penelitian terdahulu maupun teori agar penelitian yang dihasilkan bisa lebih baik lagi.

REFERENCES

- [1] R. T. Vulansari, *Data Mining Teori dan Aplikasi Rapidminer*, Yogyakarta: Gavamedia, 2017.
- [2] R. Sanjaya, *Belajar Data Mining Dengan Rapidminer*, Yogyakarta: Andi, 2015.
- [3] D. & G. W. N. Nofriansyah, *Algoritma Data Mining dan Pengujian*, Yogyakarta: Deepublish, 2015.
- [4] Adiyanto, "Prediksi Harga Crude Palm Oil Menggunakan Metode Support Vector Machine Dengan Optimasi Parameter Menggunakan Algoritma Genetika," *Ipsikom*, pp. 1-17, 2015.
- [5] E. A. T. W. & E. A. Mahmudah, "Prediksi Harga Minyak Kelapa Sawit (CPO) Menggunakan Kombinasi Metode Fuzzy-Mamdani dan Evolution Strategis.," *Telkom University*, pp. 1-7, 2015.
- [6] D. I. Puspitasari, "Penerapan Data Mining Menggunakan Perbandingan Algoritma Greedy Dengan Algoritma Genetika Pada Prediksi Rentet Waktu Harga Crude Palm Oil," *UIK MAB Banjarmasin*, pp. 1-6, 2017.
- [7] R. E. Caraka, "Peramalan Crude Palm Oil (CPO) Menggunakan Support Vector Regression Kernel Radial Basis," *Binus University*, pp. 1-15, 2017.
- [8] D. R. C. W. & R. S. P. Rahayu, "Implementasi Metode Backpropagation Untuk Klasifikasi Kenaikan Harga Minyak Kelapa Sawit," *Universitas Brawijaya*, pp. 1-6, 2018.
- [9] H. Aprina, "ANALISIS PENGARUH HARGA CRUDE PALM OIL (CPO) DUNIA TERHADAP NILAI TUKAR RIIL RUPIAH," *Buletin Ekonomi Moneter dan Perbankan*, pp. 1-24, 2015.
- [10] M. E. A. D. B. A. & N. N. Ariyanto, "Analisis Harga Minyak Sawit, Tinjauan Kointegrasi Harga Minyak Nabati dan Minyak Bumi," *Jurnal Manajemen dan Agribisnis*, pp. 1-15, 2015.
- [11] A. Halim, *Analisis Investasi*, Jakarta: Salemba Empat, 2015.
- [12] I. Fahmi, *Analisis Investasi Dalam Perspektif Ekonomi dan Politik*, Bandung: Refika Aditama, 2016.

- [13] Nugroho, SVM : Paradigma Baru Dalam SoftComputing dan Aplikasinya, Bali: Konferensi Nasional Sistem dan Informatika, 2018.
- [14] Dr.Suyanto, Machine Learning Teknik Dasar dan Lanjutan, Bandung: Informatika, 2018.
- [15] Dr.Suyanto, Data Mining Untuk Klasifikasi dan Klasterisasi Data, Bandung: Informatika, 2019.
- [16] Dr.Sudaryono, Metodologi Riset di Bidang TI, Yogyakarta: Andi, 2015.