

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Investasi Emas**

Emas merupakan unsur kimia dalam tabel periodik yang memiliki simbol Au (bahasa latin : ‘aurum’) dan nomor atom yang dimiliki emas adalah 79. Au merupakan unsur transisi dalam sistem periodik unsur, mengkilap, kuning, berat, “malleable”, dan “ductile”. Saat ini ada berbagai macam bentuk emas yang ada, yaitu:

1. Emas Batangan
2. Emas logam seperti koin emas
3. Emas perhiasan seperti cincin, anting, gelang, kalung, dan jam tangan.

Dataset yang digunakan dalam penelitian ini merupakan harga jual emas batangan persatu gram yang tersedia pada web <https://harga-emas.org/>. Emas batangan merupakan emas yang paling banyak dijadikan sebagai sarana investasi karena emas batangan memiliki tingkat kemurnian mencapai 99.99% dengan kata lain termasuk ke dalam emas murni.

Sifat dari emas batangan ini adalah tetap, likuid atau mudah dicairkan, dan mudah dijual kembali jika pemilik terdesak membutuhkan uang. Hal ini yang menjadikan emas batangan menguntungkan dan aman sebagai salah satu instrumen investasi. Investasi emas memiliki kelebihan dan kekurangan.

Investasi dapat diartikan sebagai kegiatan yang bertujuan untuk mengembangkan harta dengan tujuan untuk memperoleh keuntungan pada masa yang akan datang. Salah satunya yaitu investasi emas, selaras dengan

pedapat dari (Johari, 2017) Investasi emas baik itu berupa koin emas, emas batangan, ataupun perhiasan emas memiliki hasil yang lebih tinggi bila dibandingkan dengan deposito. Dan jauh lebih stabil bila dibandingkan dengan saham. Menurut para pakar keuangan, investasi emas berada pada posisi kedua setelah usaha dalam sektor riil yang berjalan dengan baik.

Kemudian untuk kelebihan dari investasi emas dituturkan kembali oleh (Johari, 2017) Kelebihan investasi emas yaitu bebas pajak (*tax free*) di Indonesia, karena emas batangan dimasukkan sebagai komoditi produksi yang tidak kena pajak. Sehingga dengan berinvestasi pada emas batangan, maka dapat diindikasikan telah berinvestasi pada aset yang bebas pajak.

Menurut (Budiman, 2020), Investasi merupakan salah satu cara untuk meraih laba dimasa yang akan datang. Emas merupakan logam mulia yang banyak diminati oleh investor karena harganya yang cenderung naik, selain mudah dicairkan atau memiliki tingkat likuiditas yang tinggi.

Berikut Kelebihan dan Kekurangan Investasi Emas:

- Kelebihan Investasi Emas
  1. Emas memiliki likuiditas yang tinggi, yang berarti emas mudah untuk dicairkan ke dalam bentuk uang tunai dalam waktu yang relatif singkat dan dengan cara lebih mudah dibandingkan instrumen investasi lainnya.
  2. Investasi emas terbebas dari pajak karena keuntungan didapatkan dari selisih harga saat membeli dan menjual emas tersebut.

3. Harga emas cenderung stabil, bahkan terus mengalami kenaikan setiap tahunnya sehingga saat dijual kembali akan mendapatkan imbal hasil yang tinggi.
  4. Emas melindungi investor dari inflasi, sebab jika terjadi inflasi emas akan melindungi kekayaan pemiliknya dibanding dengan investasi yang langsung berkaitan dengan uang.
- Kekurangan Investasi Emas
    1. Risiko kehilangan emas cukup tinggi, apabila emas tersebut berbentuk perhiasan yang digunakan sehari-hari. Risiko muncul dari aksi kejahatan seperti pencurian atau perampokan.
    2. Harga emas cenderung melambat ketika kondisi ekonomi negara stabil, hal ini dikarenakan saat kondisi ekonomi sedang baik-baik saja, maka akan mendorong masyarakat untuk konsumsi yang sewajarnya.
    3. Emas mempunyai risiko berfluktuasinya harga setiap hari, resiko ini disebut dengan jenis investasi data time series.
    4. Emas tidak cocok dijadikan investasi jangka pendek, karena dengan waktu yang singkat tidak memberikan keuntungan investasi yang signifikan.

Berdasarkan beberapa teori di atas, dapat disimpulkan bahwa investasi bertujuan untuk mengembangkan harta agar mendapat keuntungan di masa yang akan datang. Salah satunya Investasi emas, Investasi emas disamping lebih stabil dibandingkan saham, investasi emas bebas pajak dan harganya cenderung naik, mudah dicairkan dan likuiditasnya tinggi.

## 2.2 Data Mining

### 2.2.1 Pengertian Data Mining

*Data mining* adalah suatu proses logis yang digunakan untuk mencari dan menemukan pola melalui sejumlah besar data. Tujuan dari teknik ini adalah untuk menemukan pola yang sebelumnya tidak diketahui. Dalam *data mining* tumpukan data masa lalu ini dianggap sebagai tambang yang dapat diolah untuk menghasilkan sebuah pengetahuan yang berharga (Jakarta, 2017).

Berikut ini adalah beberapa metode data mining, diantaranya adalah :

1. *Classification*

Metode *Classification* adalah suatu tindakan untuk memberikan kelompok kepada setiap keadaan.

2. *Clustering*

Metode *Clustering* ini digunakan untuk mengidentifikasi kelompok alami dari kasus yang didasarkan kepada sebuah kelompok atribut untuk mengelompokkan data yang memiliki kemiripan.

3. *Association*

Metode *Association* bisa disebut juga sebagai *Market Basket Analysis*. Metode ini biasanya digunakan untuk mengatasi sebuah *problem* yang mengidentifikasi produk-produk yang sering dibeli secara bersamaan oleh *customer*, sehingga tujuan dari metode ini yaitu :

- a. Mencari produk yang biasanya terjual secara bersamaan
- b. Untuk mencari tahu aturan apa yang menyebabkan kesamaan tersebut.

#### 4. *Regression*

Metode *Regression* bertujuan untuk mencari sebuah pola dan menemukan sebuah nilai numerik, tetapi metode *regression* ini tidak bisa mencari pola yang dijabarkan sebagai class.

#### 5. *Forecasting*

Teknik *Forecasting* dapat membantu pertanyaan seperti “Sebanyak apa penjualan produk tertentu pada bulan depan?”. Dalam prosesnya teknik *Forecasting* akan mengambil sederetan angka dimana angka tersebut menunjukkan nilai yang berjalan seiring waktu, kemudian teknik *Forecasting* akan menghubungkan nilai masa depan dengan menggunakan berbagai macam teknik. Seperti teknik *machine-learning* dan teknik statistik yang berhubungan dengan *trend*, musim, dan *noise* pada data.

#### 6. *Sequence Analysis*

Metode *Sequence Analysis* ini digunakan untuk mencari pola pada serangkaian kejadian yang disebut dengan *Sequence*.

#### 7. *Deviation Analysis*

Metode *Deviation Analysis* penggunaannya sangat luas, biasanya metode ini digunakan untuk mencari kasus yang bertindak sangat berbeda dari normalnya.

Menurut (Saleh, 2015), Data mining dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif. Tahapnya adalah sebagai berikut:

##### 1. Pembersihan data (*Data Cleaning*)

Pembersihan data, merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.

2. Integrasi data (*Data Integration*)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam suatu database baru.

3. Seleksi data (*Data Selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

4. Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *Data Mining*.

5. Proses *Mining*

Proses *Mining* merupakan proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi pola (*Pattern Evolution*)

Mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.

7. Presentasi pengetahuan (*Knowledge Presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

Menurut pendapat (Susanto & Suryani, 2010) terdapat enam fungsi dalam data mining, yaitu :

1. Fungsi deskripsi (*description*),

2. Fungsi estimasi (*estimation*),
3. Fungsi prediksi (*prediction*),
4. Fungsi klasifikasi (*clasification*), dan
5. Fungsi asosiasi (*association*).

Menurut (Orpa et al., 2019), Prediksi atau *prediction* adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perikiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi.

### **2.3 Prediksi**

Prediksi adalah suatu proses memperkirakan sistematis mengenai sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti tentang yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi. Selain itu menurut kamus besar bahasa Indonesia, prediksi memiliki makna kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan

merupakan input bagi proses perencanaan dan pengambilan keputusan (MASKAPAI et al., 2013).

#### 2.4 Algoritma Support Vector Machine (SVM)

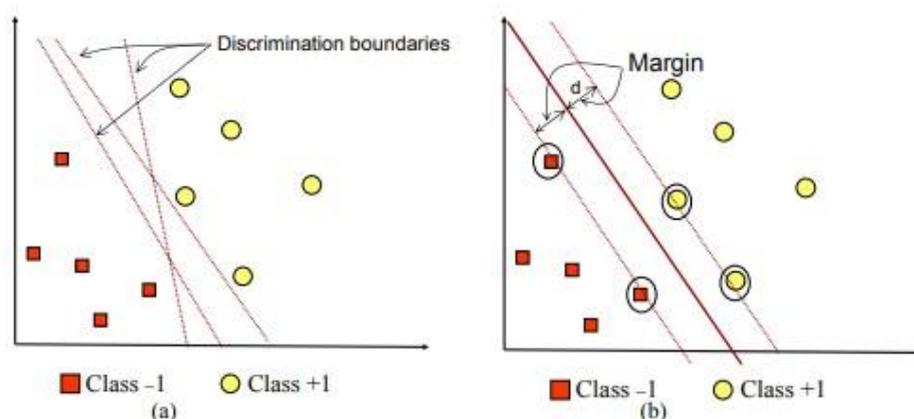
SVM dikembangkan oleh Boser, Guyon, Vapnik, dan pertama kali dipresentasikan pada tahun 1992 di Annual Workshop on Computational Learning Theory. Konsep dasar SVM adalah kombinasi dari teori-teori komputasi yang telah ada puluhan tahun sebelumnya, seperti *margin hyperplane*, *kernel*, dan konsep-konsep pendukung yang lain. Akan tetapi hingga tahun 1992, belum pernah ada upaya merangkaikan komponen-komponen tersebut.

SVM adalah metode *machine learning* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperlane* terbaik yang memisahkan dua buah class pada input space.

Prinsip dasar SVM adalah *linear classifier*, dan selanjutnya dikembangkan agar dapat bekerja pada *problem non-linear* dengan konsep kernel *trick* pada ruang kerja berdimensi tinggi. Hal ini meningkatkan minat penelitian pada bidang *pattern recognition* untuk mengetahui potensi kemampuan SVM secara teoritis maupun aplikasi.

Menurut (Perdana & Furqon, 2018), Support Vector Machine adalah algoritma *supervised* yang berupa klasifikasi dengan cara membagi data menjadi dua kelas menggunakan garis vektor yang disebut *hiperlane*.

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperlane* terbaik yang berfungsi sebagai pemisah dua buah *class* pada *input space*. Gambar 1a 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. Proses pembelajaran dalam problem klasifikasi diterjemahkan sebagai upaya menemukan garis (*hyperlane*) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) ditunjukkan pada gambar 2.1 (Nugroho, 2008).



Gambar 2.1 SVM Berusaha Hyperlane terbaik yang memisahkan kedua Class Negatif dan Positif (Nugroho, 2008)

*Hyperlane* pemisah terbaik antara kedua *class* dapat ditemukan dengan mengukur *margin hyperlane* tersebut dan mencari titik maksimalnya. *Margin* adalah jarak antara *Hyperlane* tersebut dengan data terdekat dari masing-masing *class*. Subnet data *training set* yang paling dekat ini disebut sebagai *support vector*. Garis solid pada Gambar 2.1. menunjukkan *Hyperlane* yang terbaik, yaitu yang terletak tepat

pada tengah-tengah kedua *class*, sedangkan titik kotak dan lingkaran yang berada dalam lingkaran hitam adalah *support vector*. Upaya mencari lokasi *Hyperlane* optimal ini merupakan inti dari proses pembelajaran pada SVM (Nugroho, 2008).

Data yang tersedia dinotasikan sebagai  $\vec{x} \in R^d$  sedangkan label masing-masing dinotasikan  $y_i \in \{-1 + 1\}$  untuk  $i = 1, 2, \dots, l$ , yang mana  $l$  adalah banyaknya data. Diasumsikan kedua *class* -1 dan +1 dapat terpisah secara sempurna oleh *hyperlane* berdimensi  $d$ , yang didefinisikan :

$$\vec{w} \cdot \vec{x} + b = 0 \dots\dots\dots (1)$$

$w$  = Bidang normal

$b$  = Posisi bidang relatif terhadap pusat koordinat

Sebuah *pattern*  $\vec{x}_i$  yang termasuk *class* -1 (sampel negatif) dapat dirumuskan sebagai *pattern* yang memenuhi pertidaksamaan :

$$\vec{w} \cdot \vec{x}_i + b \leq -1 \dots\dots\dots (2)$$

Sedangkan *pattern*  $\vec{x}$  yang termasuk *class* +1 (sample positif) :

$$\vec{w} \cdot \vec{x}_l + b \geq +1 \dots\dots\dots (3)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperlane* dan titik terdekatnya, yaitu  $1/\|\vec{w}\|$ . Hal ini dapat dirumuskan sebagai

Quadratic Programming (QP) problem, yaitu mencari titik minimal persamaan (4), dengan memperhatikan constraint persamaan (5).

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \dots\dots\dots (4)$$

$$y_i(\vec{x}_i \cdot \vec{w} + b) - 1 \geq 0, \forall i \dots\dots\dots (5)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, di antaranya Language Multipler sebagaimana ditunjukkan pada persamaan (6).

$$L(\vec{w}, b, \alpha) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^l \alpha_i (y_i ((\vec{x}_i \cdot \vec{w} + b) - 1)) \quad (i = 1, 2, \dots, l) \quad (6)$$

$\alpha_i$  adalah Lagrange multipliers, yang bernilai nol atau positif ( $\alpha_i \geq 0$ ). Nilai optimal dari persamaan (6) dapat dihitung dengan meminimalkan L terhadap  $\vec{w}$  dan  $b$ , dan memaksimalkan L terhadap  $\alpha_i$ . Dengan memperhatikan sifat bahwa pada titik optimal gradient  $L = 0$ , persamaan (6) dimodifikasi sebagai maksimalisasi problem yang hanya mengandung saja  $\alpha_i$ , sebagaimana persamaan (7).

Maximize:

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \vec{x}_i \vec{x}_j \quad (7)$$

Subject to:

$$\alpha_i \geq 0 \quad (i = 1, 2, \dots, l) \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (8)$$

Hasil perhitungan ini diperoleh  $\alpha_i$  yang kebanyakan bernilai positif. Data yang berkorelasi dengan  $\alpha_i$  positif inilah yang disebut sebagai *support vector*. (Nugroho, 2008).

Pada Rapidminer, operator SVM mendukung berbagai jenis *kernel\_type*, diantaranya:

- dot : *Kernel* titik didefinisikan oleh  $k(x,y)=x*y$  yaitu produk dalam dari x dan y.
- radial : *Kernel* radial didefinisikan oleh  $\exp(-g \|xy\|^2)$  dimana g adalah gamma, ditentukan oleh parameter gamma *kernel*. Parameter gamma yang dapat disesuaikan memainkan peran utama dalam kinerja *kernel*, dan harus disesuaikan dengan masalah yang dihadapi.
- Polinomial : *Kernel* polinomial didefinisikan oleh  $k(x,y)=(x*y+1)^d$  dimana d adalah derajat polinomial dan ditentukan oleh parameter derajat *kernel*. *Kernel* polinomial sangat cocok untuk masalah yang dimana semua data pelatihan di normalisasi.
- Neural : *kernel* saraf didefinisikan oleh jaringan saraf dua lapis  $\tanh(ax*y+b)$  dimana a adalah alfa dan b adalah konstanta intersep. Parameter ini dapat disesuaikan menggunakan parameter kernel a dan

*kernel* b. Nilai umum untuk alfa  $1/N$ , dimana  $N$  adalah dimensi data. Perhatikan bahwa tidak semua pilihan  $a$  dan  $b$  mengarah ke fungsi *kernel* yang valid.

- Anova : *Kernel* anova didefinisikan dengan pangkat  $d$  dari penjumlahan  $\exp(-g(xy))$  dimana  $g$  adalah gamma dan  $d$  adalah derajat. Gamma dan derajat masing-masing disesuaikan dengan parameter *kernel* gamma dan derajat *kernel*.
- Epachnenikov : adalah fungsi ini  $(3/4)(1-u^2)$  untuk  $u$  antara  $-1$  dan  $1$  dan nol untuk  $u$  diluar kisaran itu. Ini memiliki dua parameter *kernel*  $\sigma_1$  dan derajat *kernel* yang dapat disesuaikan.
- Gaussian\_combination : adalah *kernel* kombinasi gaussian. Memiliki parameter *kernel*  $\sigma_1$ , *kernel*  $\sigma_2$  dan *kernel*  $\sigma_3$  yang dapat disesuaikan.
- Multiquadric : *kernel* multiquadric didefinisikan oleh akar kuadrat dari  $\|xy\|^2+c^2$ . Ini memiliki parameter yang dapat disesuaikan  $\sigma_1$  dan *kernel*  $\sigma$  shift.

Kelebihan dan kekurangan SVM :

- Kelebihan

1. Generalisasi

Didefinisikan sebagai kemampuan untuk menghasilkan suatu *pattern*, yang tidak termasuk data yang dipakai dalam fase pembelajaran metode itu. Vapnik menjelaskan bahwa *generalization error* dipengaruhi oleh

dua faktor: *error* terhadap *training set*, dan satu faktor lagi yang dipengaruhi oleh dimensi VC (Vapnik-Chervokinensis). Strategi pembelajaran pada neural network dan umumnya learning machine difokuskan pada usaha untuk meminimalkan error pada training-set. Strategi ini disebut *Empirical Risk Minimization* (ERM). Adapun SVM selain meminimalkan error pada training-set, juga meminimalkan faktor kedua. Strategi ini disebut *Structural Risk Minimization* (SRM), dan dalam SVM diwujudkan dengan memilih *hyperlane* dengan margin terbesar. Berbagai studi empiris menunjukkan bahwa pendekatan SRM dan SVM memberikan *error* generalisasi yang lebih kecil dari pada yang diperoleh dari strategi ERM pada neural network maupun metode lain.

## 2. *Curse of dimensionality*

Didefinisikan sebagai masalah yang dihadapi suatu metode *pattern recognition* dalam mengestimasi parameter (misalnya jumlah *hidden neuron* pada neural network, *stopping criteria* dalam proses pembelajaran dsb.) dikarenakan jumlah sampel data yang relatif sedikit dibandingkan dimensional ruang vektor tersebut. Semakin tinggi dimensi dari ruang vektor informasi yang diolah, membawa konsekuensi dibutuhkannya jumlah data dalam proses pembelajaran. Pada kenyataanya seringkali terjadi, data yang diolah berjumlah terbatas, dan untuk mengumpulkan data yang lebih banyak tidak mungkin dilakukan karena kendala biaya dan kesulitan teknis. Dalam

kondisi tersebut, jika metode itu “terpaksa” harus bekerja pada data yang berjumlah relatif sedikit dibandingkan dimensinya, akan membuat proses estimasi parameter metode menjadi sangat sulit.

### 3. Landasan teori

Sebagai metode yang berbasis statistik, SVM memiliki landasan teori yang dapat dianalisa dengan jelas, dan tidak bersifat black box.

### 4. *Feasibility*

SVM dapat diimplementasikan relatif mudah, karena proses penentuan *Support Vector* dapat dirumuskan dalam *QP problem*. Dengan demikian jika kita memiliki library untuk menyelesaikan *QP problem*, dengan sendirinya SVM dapat diimplementasikan dengan mudah. Selain itu dapat diselesaikan dengan metode sekuensial sebagaimana penjelasan sebelumnya.

- Kekurangan

1. Sulit dipakai dalam problem berksala besar. Skala besar dalam hal ini dimaksudkan dengan jumlah sampel yang diolah.
2. SVM secara teoritik dikembangkan untuk problem klasifikasi dengan dua class. Dewasa ini SVM telah dimodifikasi agar dapat menyelesaikan masalah dengan class lebih dari dua, antara lain strategi *One Versus rest* dan strategi *Tree Structure*.

## 2.5 Komponen Perhitungan Prediksi dalam SVM

### 2.5.1 *Kernel*

Banyak teknik data mining yang dikembangkan dengan asumsi kelinieran. Sehingga Algoritma yang dihasilkan terbatas untuk kasus-kasus yang linier. Karena itu, bila suatu kasus klasifikasi memperlihatkan ketidaklinierean, algoritma seperti *persepctron* tidak bisa mengatasinya. Secara umum, kasus-kasus di dunia nyata adalah kasus yang tidak linier. Semisal data yang sulit dipisahkan secara linier, metode *kernel* adalah salah satu untuk mengatasinya. Dengan metode *kernel* suatu data  $x$  di *input space* dimapping ke *feature space*  $F$  dengan dimensi yang lebih tinggi melalui map  $\varphi$  sebagai berikut  $\varphi : x \rightarrow \varphi(x)$ . Karena itu data  $x$  di *input space* menjadi  $\varphi(x)$  di *feature space*.

Fungsi kernel yang harus digunakan untuk substitusi *dot product* di *feature space* sangat bergantung pada data. Biasanya metode *cross-validation* digunakan untuk pemilihan fungsi *kernel* ini. Pemilihan fungsi *kernel* yang tepat adalah hal yang sangat penting. Karena fungsi *kernel* ini akan menentukan *feature space* untuk fungsi prediksi yang akan dicari (Pudjianto et al., 2015).

### 2.5.2 $C$ (*cost*)

Menurut (Pudjianto et al., 2015), pada perhitungan SVM diberikan penalti dengan menambahkan nilai *cost*  $C$ . Nilai *cost*  $C$  dipilih untuk mengontrol keseimbangan antara nilai *margin* dan *error* prediksi. Semakin besar nilai  $C$ , maka penalti yang diberikan terhadap data *error* juga semakin besar.

### 2.5.3 RMSE (*Root Mean Square Error*)

Menurut (Pudjianto et al., 2015), keakuratan keseluruhan dari setiap model peramalan naik itu rata-rata bergerak, *eksponensial smoothing* atau lainnya dapat dijelaskan dengan membandingkan nilai yang diproyeksikan dengan nilai aktual atau nilai yang diamati.

Menurut (Anggie, 2018), perhitungan *error* merupakan pengukuran kesalahan antara *output* aktual dan *output* target. Langkah yang digunakan adalah *Sum Square Error* yang merupakan hasil penjumlahan nilai kuadrat *error* neuron1 dan neuron2 pada setiap lapisan *output* pada setiap data. Hasil penjumlahan keseluruhan nilai *Sum Square Error* akan digunakan dalam menghitung nilai *Root Mean Square Error* (RMSE) tiap iterasi. Nilai RMSE yang kecil menunjukkan hasil akurasi prediksi yang terbaik, yang dipersentasikan dengan persamaan.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (X_t - F_t)^2}{n}} \dots\dots\dots (9)$$

Dengan keterangan sebagai berikut :

- X<sub>t</sub>           = Nilai aktual pada periode ke t
- F<sub>t</sub>           = Nilai peramalan pada periode ke t
- X<sub>t</sub>-F<sub>t</sub>       = Nilai kesalahan (error) pada periode ke t
- N             = Jumlah data

## 2.6 Regresi Linear

Menurut (Katemba & Djoh, 2017), metode regresi linear adalah alat statistik yang dipergunakan untuk mengetahui pengaruh antara satu atau beberapa variabel terhadap satu buah variabel. Kelebihan atau manfaat dari regresi linear diantaranya analisis regresi lebih akurat dalam melakukan analisis korelasi, karena analisis itu kesulitan dalam menunjukkan tingkat perubahan suatu variabel terhadap variabel lainnya (slop) dapat ditentukan. Dengan analisis regresi peramalan atau perkiraan nilai variabel terikat pada nilai variabel bebas lebih akurat.

Menurut (Anggie, 2018), metode regresi merupakan salah satu teknik analisis statistika yang digunakan untuk menggambarkan hubungan antara satu variabel respon dengan satu atau lebih variabel penjelas. Metode regresi terdapat 2 macam yaitu : Regresi Linear dan Regresi Non Linear. Regresi Linear mempunyai model dengan 1 variabel bebas dan model dengan >1 variabel bebas (regresi linear berganda). Sedangkan regresi non linear mempunyai model persamaan eksponensial (In) dan model persamaan berpangkat (log). Dalam metode regresi linear sederhana mempunyai data yang nantinya digunakan sebagai bahan untuk membentuk persamaan regresi. Persamaan regresi yaitu :

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (10)$$

Dengan keterangan sebagai berikut :

- Y           = variabel dependen
- a           = konstanta
- b           = koefisien variabel x

$x$  = variabel independen

Konstanta  $a$  dan  $b$  diperoleh dari persamaan (11) (12), dimana nilai  $x$  dan  $y$  diperoleh dari data-data sebelumnya yang dijadikan dalam bentuk sampel (Anggie, 2018).

$$a = \bar{y} - b\bar{x} \quad (11)$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (12)$$

Metode Regresi Linear sebagai metode statistik seringkali digunakan dalam melakukan prediksi. Pengerjaanya yang efisien dengan nilai akurasi yang cukup tinggi merupakan nilai lebih yang dimiliki metode ini (Anggie, 2018).

## 2.7 Rapidminer

Rapidminer yang dapat dijalankan pada berbagai sistem operasi dan bersifat terbuka (*open source*). Rapidminer dapat menjadi sebuah solusi terhadap *data mining*, *text mining*, dan analisis prediksi. Rapidminer menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Rapidminer juga menempati peringkat pertama sebagai *software* data mining pada polling Kdnuggets, sebuah portal *data-mining* pada 2010-2011.

Rapidminer memiliki beberapa fitur, diantaranya adalah :

1. Banyaknya variasi plugin, seperti *text plugin* untuk melakukan *analisis text*.

2. Memiliki kurang lebih 500 *operator* data mining. Termasuk *operator input*, *output*, *data preprocessing* dan visualisasi.
3. Bentuk grafis yang canggih, seperti *tree chart* dan *3d scatter plots*.
4. Mengintegrasikan proyek data mining Weka dan statistika R.
5. Operator-operator dalam proses data mining tersusun atas operator-operator yang *nestable*, dideskripsikan dengan XML, dan dibuat dengan GUI.
6. Menyediakan *machine learning* dan prosedur data mining.

## **2.8 State Of The Art**

### **2.8.1 Penelitian Terdahulu**

Penelitian terdahulu sangat berperan penting pada penelitian, dimana dengan adanya penelitian terdahulu sangat berpengaruh dalam memperkaya teori dan referensi wawasan pengetahuan bagi penelitian ini. Ulasan dari penelitian terkait, dilakukan dengan maksud dan tujuan untuk menganalisis penelitian yang telah dilakukan sebelumnya. Penelitian terdahulu dapat dilihat sebagai judul berikut:

#### **1. Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko**

Nama penulis dari jurnal tersebut adalah Mohammad Guntur, Julius Santony, dan Yuhandri dengan tahun terbitan jurnal yaitu Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) Vol. 2 No. 1 (2018) 354 – 360.

Penelitian yang dilakukan pada jurnal tersebut adalah memprediksi status harga emas dimasa depan, karna harga emas berfluktuasi setiap hari atau disebut juga dengan jenis investasi *data time series*. Maka untuk menghindari atau meminimalisir risiko dari investasi emas di masa depan penelitian ini membuat prediksi harga emas dengan menggunakan Algoritma Naïve Bayes Classifier sebagai metode untuk merancang perhitungan harga emas dimasa yang akan datang secara akurat. Data yang digunakan didalam penelitian ini diantaranya bersumber dari harga logam mulia emas Pegadaian, Nilai tukar rupiah terhadap mata uang Dollar AS, Nilai tukar rupiah terhadap mata uang Euro, dan Harga minyak mentah dunia. Kemudian diimplementasikan ke dalam aplikasi Rapidminer agar dapat dijadikan solusi memprediksi harga emas di masa yang akan datang. Hasil yang didapatkan dari penelitian ini, dari 16 data yang diuji, terdapat 12 data yang hasil ujinya sama dengan data riil dengan nilai akurasi mencapai 75%. Dan kesimpulanya Algoritma Naïve Bayes Classifier mampu melakukan prediksi harga emas dengan baik.

## **2. Prediksi Harga Emas Menggunakan Metode *Fuzzy Time Series Model* Algoritma Chen**

Nama penulis dari jurnal tersebut adalah Darwin Purba Sugumonrong, Aldrick Handinata, dan Anton Teja dengan tahun terbitan jurnal yaitu *INFORMATICS ENGINEERING RESEARCH AND TECHNOLOGY [VOLUME 1 NO.1 AGUS 2019]*.

Penelitian yang dilakukan pada jurnal tersebut adalah memprediksi harga emas, untuk meminimalisir kerugian yang diperoleh oleh investor. Penelitian ini menggunakan metode *fuzzy time series* untuk memprediksi harga kenaikan atau penurunan harga emas. Dasar konsep perhitungannya, penelitian ini menggunakan *fuzzy set*. Nilai prediksi harga emas model Chen digunakan dalam penelitian ini dengan tujuan untuk pengujian keakuratan nilai prediksi dengan menguji dan membandingkan hasil nilai prediksi dengan nilai kebenaran harga emas yang sebenarnya. Sumber data yang digunakan dalam penelitian ini menggunakan data sekunder yang diperoleh dari [www.harga-emas.org](http://www.harga-emas.org). Data harga emas dengan nilai kurs Rupiah ini terdiri dari periode 1 Januari 2015 hingga 31 Desember 2017, data yang diambil berupa data harian. *Software* yang digunakan dalam penelitian ini yaitu MATLAB, tujuannya untuk mengolah prediksi training data. Dan hasil dari penelitian ini untuk rata-rata data aktual dengan nilai prediksi tidak lebih dari Rp. 2.850, oleh karena itu sistem prediksi ini dapat menjadi acuan dalam memprediksi harga emas.

### **3. Prediksi Harga Emas Menggunakan Metode Neural Network Backpropagation Algoritma Conjugate Gradient**

Nama penulis dari jurnal tersebut adalah Yuslena Sari dengan tahun terbitan jurnal yaitu Jurnal ELTIKOM, Vol. 1 No.2, Desember 2017, Hal 64-70.

Penelitian yang dilakukan pada jurnal tersebut adalah memprediksi harga emas menggunakan *Artificial Neural Network Backpropagation* tujuannya yaitu untuk memprediksi harga emas yang sangat populer di decade terakhir ini.

Metode yang digunakan untuk pelatihan ini adalah *Backpropagation Artificial Neural Network* dengan optimasi pelatihan *Conjugate Gradient learning* dan di implementasikan dengan Matlab. Sistem dikembangkan berdasarkan *time series* dengan data harian dari harga emas. Dan hasilnya untuk penerapan algoritma *Conjugate Gradient Artificial Neural Network Backpropagation* dapat memprediksi dengan baik, menghasilkan nilai  $MSE=0.0313651$  untuk data *testing* dan  $MSE=0.0198012$  untuk data uji dengan settingan 100 *epoch*, *goal* = 0.1 dan *learning rate* = 0.01.

#### **4. Analisis dan Perbandingan Algoritma Data Mining Dalam Prediksi Harga Saham GGRM**

Nama penulis dari jurnal tersebut adalah Reza Maulana, dan Devy Kumalasari dengan tahun terbitan jurnal yaitu Jurnal Informatika Kaputama (JIK), Vol 3 No 1, Januari 2019

Penelitian yang dilakukan pada jurnal tersebut adalah membandingkan beberapa model algoritma tujuannya untuk menemukan algoritma yang paling akurat dalam memprediksi harga saham sehingga dapat membantu para investor untuk melihat propek investasi di masa yang akan datang. Aplikasi yang digunakan dalam penelitian ini yaitu Rapidminer. Beberapa model Algoritma yang dibandingkan dalam penelitian ini diantaranya yaitu *Neural Network*, *Linear Regression*, *Support Vector Machine*, *Gaussian Process*, dan *Polynomial Regression*. Mengukur tingkat akurasi dari tiap model algoritma menggunakan model validasi 10 *Flod Cross Validation* dan evaluasi meggunakan RMSE. Hasil

pada penelitian ini menyimpulkan bahwa data harga saham GGRM dapat diprediksi dengan menggunakan model *Neural Network*, dengan hasil akurasi prediksi RMSE 612.474 +/- 89.402 (mikro: 618.916 +/- 0.000) paling kecil dibandingkan dengan model algoritma lainnya.

#### **5. Prediksi Indeks Harga Saham Gabungan (IHSG) Menggunakan Algoritma *Neural Network***

Nama penulis dari jurnal tersebut adalah Imam Halimi, dan Wahyu Andhyka Kusuma dengan tahun terbitan jurnal yaitu JEPIN (Jurnal Edukasi dan Penelitian Informatika), Vol. 4, No.1, Juni 2018.

Penelitian yang dilakukan pada jurnal tersebut adalah memprediksi kenaikan dan penurunan IHSG, metode yang digunakan pada penelitian ini adalah *Neural Network* dan *Linear Regression*, untuk aplikasi yang digunakan pada penelitian ini adalah Rapidminer Studio 8.1 yang berfungsi untuk menganalisa data. Hasil yang didapat dari proses Rapidminer menggunakan algoritma *Neural Network* adalah berupa RMSE. Hasil penelitian dilakukan dengan menggunakan *sliding window validation*. Validasi akan dilakukan pada metode algoritma *neural network* dan *linear regression*. Uji validasi *T-Test* digunakan dalam penelitian ini untuk melakukan validasi hasil perbandingan dalam membandingkan antara algoritma *neural network* dan *linear regression*, dengan hasil akhir dapat diketahui apakah perbedaan hasil yang didapat signifikan atau tidak. Hasil pengujian *T-Test* yang didapat dari penelitian ini adalah pada pengujian hasil percobaan yang menggunakan fitur *windowing* didapat hasil

sebesar 1,000 dan untuk pengujian T-Test tanpa menggunakan *windowing* adalah sebesar 1,000. Hal ini menunjukkan bahwa tidak signifikannya perbedaan antara hasil yang didapat oleh algoritma *neural network* dan *linear regression*. Karena nilai yang didapat diatas alpha *T-Test* yaitu 0,050.

## **6. Prediksi Harga Emas Menggunakan *Fuzzy Time Series Markov Chain***

Nama penulis dari jurnal tersebut adalah Nurmalia Rukhansah, Much Aziz Muslim, dan Riza Arifudin dengan tahun terbitan jurnal yaitu KOMPUTAKI Vol. 1 No. 1 Februari 2016.

Pada penelitian ini dalam memprediksi harga emas, metode yang digunakan yaitu metode *Fuzzy Time Series Markov Chain*. *Software* Matlab R2009a dengan berdasarkan output aplikasi peramalan harga emas menggunakan metode *Fuzzy Time Series Markov Chain*. Berdasarkan metode *Fuzzy Time Series Markov Chain* dengan menggunakan *Software* Matlab R2009a diperoleh peramalan harga emas untuk periode 7 hari, yaitu 5 Maret 2015 s/d 11 Maret 2015. Kemudian untuk keakuratan dari hasil kerja aplikasi peramalan harga emas menggunakan metode *Fuzzy Time Series Markov Chain* sebesar 99,66574%. Dengan kesalahan model peramalan yang dihitung menggunakan parameter AFER dan MAE masing-masing sebesar 0,33426% dan Rp. 1.679, yang dapat disimpulkan bahwa model peramalan menggunakan *Fuzzy Time Series Markov Chain* sebesar 0,33426% dari data aktual.

## **7. Implementasi Metode K-Nearest Neighbor dan Regresi Linear dalam Prediksi Harga Emas**

Nama penulis dari jurnal tersebut adalah Prabowo Budi Utomo, Ema Utami, dan Suwanto Raharjo dengan tahun terbitan jurnal yaitu Jurnal Informasi Interaktif Vol. 4 No. 3 September 2019.

Penelitian yang dilakukan pada jurnal tersebut adalah mengimplementasikan metode K-NN dan Regresi Linear dalam memprediksi harga emas. Proses prediksi diawali dengan melakukan pembacaan pada data *training* yang telah dipersiapkan. Data yang digunakan dalam penelitian ini merupakan data yang telah dikumpulkan pada periode 4 Januari 2000 – 30 Desember 2016. Setelah data *training* memenuhi persyaratan, maka dilanjutkan dengan pemrosesan menggunakan algoritma K-NN. Hasil dari pemrosesan algoritma K-NN menjadi dasar dalam proses prediksi menggunakan algoritma Regresi Linear untuk menghasilkan keluaran yang diharapkan. Model yang didapatkan dari proses prediksi kemudian diuji menggunakan data uji yang telah dipersiapkan untuk mendapatkan nilai RMSE yang semakin kecil menentukan tingkat keberhasilan dari prediksi yang dilakukan. Hasil dari nilai parameter K sebesar 2 diperoleh nilai RMSE terendah yaitu 1.29292 dan nilai *relative error* sebesar 9,47%. Nilai ini tergolong masih besar maka hasil dari metode K-NN akan diproses dengan metode regresi linear agar nilai bisa menjadi lebih rendah. Hasil dari pengujian K-NN dan Regresi Linear menghasilkan nilai RMSE sebesar 0.05807 dan nilai *relative error* sebesar 0,39%, dimana nilai ini lebih rendah dari nilai yang diperoleh pada pengujian menggunakan metode K-NN.

## 8. Penerapan Algoritma Support Vector Machine untuk Prediksi Harga Emas

Nama penulis dari jurnal tersebut adalah Nugroho Dwi S, dengan tahun terbitan jurnal yaitu Jurnal Informatika UPGRIS Volume 1 Edisi Juni 2015.

Penelitian yang dilakukan pada jurnal tersebut adalah memprediksi harga emas dari investasi *trading* emas *online* dengan metode *Support vector machines* (SVM). Data pada penelitian ini diambil dari PT. Rifan Financindo Berjangka yang bergerak di bidang Investasi *Trading* Emas. Dataset yang digunakan merupakan data rentet waktu atau *time series*. Sedangkan untuk atribut *factory news* didapat dari kalender ekonomi dan *forex factory*. Metode dalam penelitian ini membandingkan antara dataset variabel A yang terdiri dari atribut *open*, *high*, dan *low*. Dengan dataset variabel B yang terdiri dari atribut *open*, *high*, *low*, dan *factory news*. Dalam penelitian ini aplikasi Rapidminer digunakan untuk memproses data inputan kemudian data inputan diproses di dalam aplikasi Rapidminer dengan model algoritma SVM dengan *10-foldcross validation*. Hasil yang didapatkan RMSE variabel A sebesar 4,695 dan RMSE variabel B sebesar 4,620. Dengan demikian dari hasil pengujian tersebut dapat disimpulkan bahwa variabel B lebih akurat dan memiliki nilai RMSE terendah.

## 9. Implementasi Metode Naïve Bayes Untuk Prediksi Harga Emas

Nama penulis dari jurnal tersebut adalah Frandy Ristiano, Nurmalasari, dan Ani Yoraeni dengan tahun terbitan jurnal yaitu Computer Science (CO-SCIENCE) Volume 1 No. 1 Januari 2021 | E-ISSN: 2774-9711.

Penelitian yang dilakukan pada jurnal tersebut adalah memprediksi harga emas. Data yang digunakan dalam penelitian ini bersumber dari data kurs rupiah dan harga emas pada situs badan pusat statistik pada bulan Januari 2016 sampai dengan bulan Desember 2019, kemudian data yang didapatkan dianalisa menggunakan metode klasifikasi naïve bayes yang bertujuan untuk menentukan profit “ya” dan “tidak”. Setelah itu data hasil pengolahan data dianalisis dengan menggunakan aplikasi data mining yaitu rapidminer. Aplikasi rapidminer digunakan untuk menganalisa dan mengukur atau menentukan *profit* dalam investasi emas. Kemudian mengukur tingkat akurasi dari hasil pengolahan data tersebut untuk menentukan *profit* dalam 1 tahun kedepan. Hasil akurasi yang didapat dari metode naïve bayes dengan menguji *data mining* dan *data testing* yaitu 95,92% yang berarti sudah cukup baik.

**10. *Data Mining Predictive Modeling for Prediction of Gold Prices Based on Dollar Exchange Rates, Bi Rates and World Crude Oil Prices***

Nama penulis dari jurnal tersebut adalah Iman Priyadi, Julius Santony, dan Jufriadif Na'am dengan tahun terbitan jurnal yaitu *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIMD)* Vol 2, No.2, September 2019, pp. 93 – 100 p-ISSN: 2614-3372 | e-ISSN: 2614-6150.

Penelitian yang dilakukan pada jurnal tersebut adalah memprediksi harga emas dalam *platform Gold Trading* untuk dapat memaksimalkan keuntungan yang diperoleh dan meminimalkan potensi kerugian yang ada. Data yang diolah dalam penelitian ini selain harga emas itu sendiri meliputi harga penutupan,

harga pembukaan, harga tertinggi, harga terendah dari nilai tukar dollar terhadap rupiah, harga minyak mentah dunia dan BI Rate. Data yang digunakan dalam penelitian ini diambil dari *website investing.com*. Teknik *Data Mining* yang digunakan dalam penelitian ini menggunakan fungsi Regresi Linear dengan 2 penerapan yaitu, perhitungan manual dan tahapan menggunakan Aplikasi Rapidminer. Hasil uji prediksi pada penelitian ini ada tujuh hari kerja pada bulan Februari 2019 yaitu pada tanggal 12, 13, 14, 15, 18, 19, dan 20. Kemudian hasil dari penereapan regresi linear di implementasikan pada Metatrader. Dan hasil yang tercapai sesuai prediksi diantara tanggal 12, 13, 14, 15, 18, 19, dan 20 ada 6 hari yang sesuai prediksi, sehingga tingkat keberhasilan dari penelitian ini dalam memprediksi harga emas ada di persentase 85%.

Berdasarkan peneltian terdahulu, terdapat persamaan antara penelitian sebelumnya dengan penelitian yang akan dilakukan, yaitu mengenai prediksi harga eamas menggunakan algoritma SVM dan metode Regresi Linear. Adapun keterbaruan dari penelitian ini yang berbeda dengan penelitian yang sebelumnya yaitu ada pada dataset, dataset yang digunakan pada penelitian ini menggunakan dataset harga emas ketika range waktu dari mulai pada tanggal 1 januari 2016 sampai dengan 31 Desember 2020. Kemudian dari algortima yang digunakan juga berbeda, algortima yang digunakan dalam penrlitian ini yaitu metode SVM untuk memproses prediksi agar menghasilkan RMSE.

Tabel 2.1 Matriks Penelitian

No.	Peneliti	Metode								Tools		Objek Penelitian		
		Naïve Bayes	Fuzzy Time Series	Neural Network	Linear Regression	Support Vector Machine	Gaussian Process	Polynomial Regression	K-Nearest Neighbor	Rapidminer	Matlab	Emas	Saham	Batubara
1.	Mohammad Guntur, dkk (2018)	✓								✓		✓		
2.	Darwin Purba Sugumonrong, dkk (2019)		✓								✓	✓		
3.	Yuslena Sari (2017)			✓							✓	✓		
4.	Reza Maulana, dkk (2019)			✓	✓	✓	✓	✓		✓			✓	
5.	Imam Halimi, dkk (2018)			✓						✓			✓	

6.	Nurmalia Rukhansah, dkk (2016)		✓							✓	✓		
7.	Prabowo Budi Utomo, dkk (2019)			✓				✓	✓		✓		
8.	Nugroho Dwi S (2015)				✓				✓		✓		
9.	Frandy Ristiano, dkk (2021)	✓							✓		✓		
10	Eko Pudjianto, dkk (2015)				✓				✓		✓		✓
11.	Iman Priyadi, dkk (2019)			✓					✓		✓		
12.	Dhafi Yarist Kusnaedi (2021)			✓	✓				✓		✓		