



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Saham**

Saham adalah satuan nilai atau pembukuan dalam berbagai instrumen finansial yang mengacu pada bagian kepemilikan sebuah perusahaan (BEI, 2017). Dengan menerbitkan saham, memungkinkan pendanaan jangka panjang untuk menjual kepentingan dalam bisnis dengan imbalan uang tunai. Pergerakan harga saham dapat dipengaruhi oleh berbagai faktor antara lain suku bunga, psikologi investor, faktor politik, dan bencana alam (Sia dan Hansun, 2014).

Jika suku bunga kemungkinan akan turun, saham akan mudah dilirik pembeli dan harga pun akan meningkat. Tetapi kemudian ada juga tingkat permintaan untuk barang-barang dalam ekonomi, yang akan mendorong naik harga saham ketika jumlah permintaan tersebut meningkat, dan tentu bisa terjadi sebaliknya. Tetapi untuk permintaan barang-barang dari luar negeri akan bergantung pada nilai tukar mata uang, dimana sebagian dari nilai mata uang juga tergantung pada tingkat suku bunga dan sebagainya. Oleh karena itu, harga saham sangat bergantung terhadap suku bunga (Sia dan Hansun, 2014).

Psikologi investor pada umumnya banyak orang mungkin akan terjun ke dalam pasar yang sudah overheated, dimana sebetulnya lebih baik tidak memasuki pasar saat itu. Psikologi investor juga pada umumnya, mereka panik dan kabur dari pasar ketika sebenarnya itulah saat yang tepat untuk melakukan aksi beli. Di dalam sejarah perekonomian, telah dilihat bagaimana pasar bisa melampaui dan

mendorong harga naik ke tingkat yang tidak sewajarnya jika dibandingkan dengan kinerja ekonomi yang nyata. Ada juga hal sebaliknya yang terjadi di pasar saham, dimana orang-orang melakukan aksi jual saat ekonomi sedang baik-baik saja, hanya karena sentiman sedang negatif (Sia dan Hansun, 2014).

Faktor politik dan bencana alam. Adanya pemilihan umum, pembunuhan tokoh penting, serangan teroris, bencana alam, *epidemic*, wabah, dan banyak kejadian lainnya dimana bisa memberikan kejutan yang terjadi besok dan hilang di hari selanjutnya, atau akan ada selama waktu bertahun-tahun, yang bisa membuat nilai saham naik atau turun. Perlu dicatat bahwa ini juga faktor non-ekonomi, yang artinya pasar juga merefleksikan hal ini. Beberapa faktor bisa memicu harga saham naik, sementara beberapa faktor lainnya bisa memicu penurunan harga (Sia dan Hansun, 2014).

## 2.2 Regresi

Menurut Suyanto (2018), regresi (*regression*) merupakan proses identifikasi relasi dan pengaruhnya pada nilai-nilai objek. Regresi memiliki tujuan, yakni menemukan suatu fungsi yang memodelkan data dengan meminimalkan galat atau selisih antara nilai prediksi dengan nilai sebenarnya. Regresi dapat dilihat sebagai alat ukur untuk mengetahui adanya korelasi antar variabel dan bisa juga untuk menentukan tingkat perubahan suatu variabel terhadap variabel lainnya (Suyanto, 2018).

Umumnya regresi digunakan untuk prediksi (*prediction*) dan peramalan (*forecasting*). Perbedaan prediksi digunakan untuk memperkirakan nilai-nilai

data bertipe apa saja dan kapan saja (masa lalu, sekarang, dan masa depan). Sedangkan, peramalan digunakan untuk memperkirakan nilai-nilai data *time series* di masa depan (Suyanto, 2018).

### 2.3 Python

Python adalah bahasa pemrograman model skrip (*scripting language*) yang berorientasi objek. Python dapat digunakan untuk pelbagai keperluan pengembangan perangkat lunak dan dapat berjalan di banyak sistem operasi. Python merupakan bahasa pemrograman yang *freeware* atau perangkat lunak bebas dalam arti sebenarnya, tidak ada batasan dalam penyalinannya atau mendistribusikannya. Lengkap dengan *source code*, *debugger*, dan *profiler*, antarmuka yang terdapat di dalam untuk layanan antarmuka, fungsi sistem, *Graphic User Interface* (GUI), dan basis datanya (Beazley, 2006). Fitur-fitur yang dimiliki Python, diantaranya (Beazley, 2006).

1. Memiliki *libraries* yang luas, dalam distribusi Python telah disediakan modul-modul.
2. Memiliki tata bahasa yang jelas dan mudah dipelajari.
3. Memiliki aturan layout *source code* yang memudahkan pengecekan, pembacaan kembali dan penulisan ulang *source code*.
4. Berorientasi objek.
5. Memiliki sistem pengelolaan memori otomatis (*garbage collection*) seperti Java, sehingga dapat mengatur penggunaan memori komputer secara otomatis

dan para *developer* tidak perlu melakukan pengaturan ingatan komputer secara langsung.

6. Modular, mudah dikembangkan dengan membuat modul-modul baru. Modul-modul tersebut dapat dibangun dengan bahasa Python atau pun C/C++.

## 2.4 Flask

Merupakan suatu *web application microframework* yang terdapat pada bahasa pemrograman Python dan berdasarkan Werkzeug *toolkit* dan *template* Jinja2. (Rizkanda, dkk., 2017). Disebut *microframework* karena Flask tidak membutuhkan *tools* atau *libraries* tertentu. Flask tidak memiliki *database abstraction layer*, *form validation*, atau komponen lain di mana sudah ada *libraries* pihak ketiga yang menyediakan fungsi umum. Namun, Flask mendukung ekstensi yang dapat menambahkan fitur aplikasi seolah-olah sudah ada di dalam Flask itu sendiri. Flask merupakan salah satu *web frameworks* Python karena ringan dan dapat mempercepat pengembangan aplikasi web sederhana. Jadi, Flask lebih cocok untuk aplikasi kecil atau aplikasi yang sederhana (Tutorials Point, 2019).

## 2.5 Penggalian Data

Menurut Prasetyo (2012), penggalian data (*data mining*) merupakan sebuah proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru

yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah *data mining* kadang disebut juga *knowledge discovery*

Salah satu teknik yang dibuat dalam *data mining* adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. Kebutuhan untuk prediksi juga dapat memanfaatkan teknik ini. Pengelompokan data juga bisa dilakukan dalam *data mining*. Pengelompokan ini bertujuan agar kita dapat mengetahui pola universal data-data yang ada. Anomali data transaksi juga perlu dideteksi untuk dapat mengetahui tindak lanjut berikutnya yang dapat diambil. Semua hal tersebut bertujuan mendukung kegiatan operasional perusahaan sehingga apa yang diharapkan bisa tercapai (Prasetyo, 2012).

Pekerjaan yang berkaitan dengan *data mining* dapat dibagi menjadi empat kelompok, yaitu model prediksi (*prediction model*), analisis kelompok (*cluster analysis*), analisis asosiasi (*association analysis*), dan deteksi anomali (*anomaly detection*) (Prasetyo, 2012). Model prediksi berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat. Ada dua jenis model prediksi, yaitu klasifikasi dan regresi. Klasifikasi digunakan untuk variabel target diskret, sedangkan regresi untuk variabel target kontinu (Prasetyo, 2012).

## 2.6 k-Neighbors Regressor (kNR)

kNR merupakan bagian dari *k-Nearest Neighbors* dengan metode regresi. Masalah regresi fokus dengan memprediksi hasil keluaran berdasarkan variabel yang diberikan dari sebuah set variabel independen. kNR membuat prediksi berdasarkan hasil dari tetangga *k* yang paling dekat dengan titiknya. Maka dari itu, untuk membuat prediksi dengan kNR terlebih dahulu harus mendefinisikan metrik untuk mengukur jarak antara titik permintaan (*query point*) dan *case* dari contoh sampel. Salah satu pilihan untuk mengukur jarak pada variabel target kontinu adalah Euclidean. Pengukuran lain termasuk Minkowski, City-block, dan Chebychev (Imandoust dan Bolandraftar, 2013). Setelah memilih nilai dari *k*, prediksi dapat dibuat berdasarkan contoh-contoh kNR. Persamaan 2.1 merupakan persamaan dari Euclidean *distance*.

$$D(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad \dots(2.1)$$

Keterangan rumus:

$D(p, q)$  = jarak antara titik  $p$  dan  $q$ .

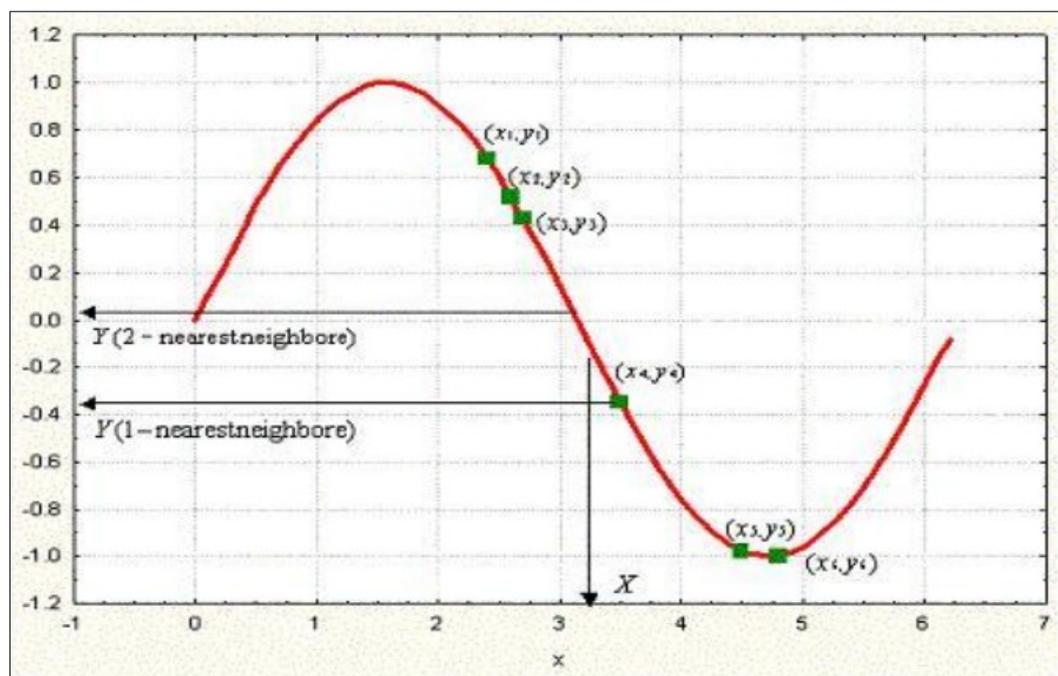
$n$  = jumlah data.

$i$  = indeks data ke- $i$ .

kNR dipertimbangkan sebagai *lazy learning* karena tidak dibuat model atau fungsi sebelumnya, tetapi merupakan hasil dari *k records* terdekat dari *training data set* yang memiliki kemiripan paling tinggi dengan tes (contoh: *query record*). Kemudian, *vote* yang menang dilakukan di antara *k records* terpilih untuk menentukan *class label* dan kemudian ditugaskan ke *query record* (Alkhatib, K.,

dkk., 2013). Perhitungan prediksi penutupan harga saham menggunakan kNR sebagai berikut.

1. Tentukan  $k$ , jumlah tetangga terdekat.
2. Hitung jarak antara sampel *training* dan *query record*.
3. Urutkan seluruh *training records* berdasarkan *distance values*.
4. Gunakan *majority vote* untuk *class labels* dari k-Nearest Neighbors dan tetapkan itu sebagai *prediction value* dari *query record*.



Gambar 2.1 Gambar Penjelasan Algoritma KNR ( Imandoust Dan Bolandraftar, 2013)

Sebagai contoh pada Gambar 2.1, di mana sebuah set titik (kotak hijau) di gambar dari hubungan antara variabel independen  $x$  dan variabel dependen  $y$  (kurva merah). Diberikan sebuah set objek hijau (disebut sebagai contoh), kemudian gunakan metode kNR untuk memprediksi hasil dari  $X$  (disebut sebagai *query point* diberikan contoh set (kotak hijau). Untuk memulainya, pertimbangkan

metode *nearest neighbor* sebagai contoh. Dalam kasus ini, mencari contoh set (kotak hijau) dan menentukan satu posisi terdekat dengan *query point* X. Untuk kasus tertentu, ini terjadi  $x_4$ . Hasil dari  $x_4$  (yaitu,  $y_4$ ) dengan demikian dianggap sebagai jawaban untuk hasil X (yaitu, Y). Jadi untuk 1- *nearest neighbor* dapat kita tulis:  $Y = y_4$ . Persamaan 2.2 merupakan rumus untuk menghitung nilai y (Imandoust dan Bolandraftar, 2013).

$$y = \frac{1}{k} \sum_{i=1}^k y_i \quad \dots(2.2)$$

Keterangan rumus:

$y$  = nilai yang ingin diprediksi.

$k$  = jumlah tetangga.

$y_i$  = nilai harga saham indeks ke-i.

Set dari *weights* (W), satu untuk setiap tetangga yang terdekat, ditentukan oleh kedekatan relatif masing-masing tetangga sehubungan dengan *query point*.

Maka seperti Persamaan 2.3:

$$W(x, p_i) = \frac{\exp(-D(x, p_i))}{\sum_{i=1}^k \exp(-D(x, p_i))} \quad \dots(2.3)$$

Keterangan rumus:

$W(x, p_i)$  = nilai bobot untuk titik  $x$  ke titik  $p_i$ .

Dengan  $D(x, p_i)$  adalah jarak antara *query point*  $x$  dan  $p_i$  ke-i adalah contoh sampel.

Bobot yang ditentukan dengan Persamaan 2.3 akan memenuhi Persamaan 2.4:

$$y = \sum_{i=1}^k W(X_0, X_i) y_i \quad \dots(2.4)$$

Keterangan rumus:

$y$  = nilai yang hendak diprediksi

$W(X_0, X_i)$  = bobot nilai dari indeks ke-0 hingga indeks ke-i

## 2.7 Root Mean Square Error (RMSE)

Merupakan metode yang berdasarkan pengukuran terhadap perbandingan antara nilai estimasi dan nilai yang sebenarnya (Shekhar dan Xiong, 2008) seperti Persamaan 2.5.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{obs,i} - X_{model,i})^2}{n}} \quad \dots(2.5)$$

Keterangan rumus:

$n$  = jumlah data.

$i$  = indeks ke-i

$X_{obs,i}$  = nilai yang diprediksi indeks ke-i

$X_{model,i}$  = nilai yang sebenarnya indeks ke-i

Persamaan 2.5 bisa dijabarkan sebagai perbandingan antara nilai *observed* (aktual),  $X_{obs}$  dan nilai estimasi,  $X_{Model}$ , yang kemudian diakarkuadratkan dan dibagi dengan jumlah dari data set ( $n$ ). Nilai *observed* sama dengan nilai yang sebenarnya dan setelah itu diakarkuadratkan. Pengukuran *error* universal ini akan digunakan untuk mengukur akurasi dari algoritma kNR.

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A