

Data Mining Menggunakan Regresi Linear untuk Prediksi Harga Saham Perusahaan Pelayaran

(Data Mining Using Linear Regression to Predict the Stock Price of Shipping Companies)

Ekka Pujo Ariesanto Akhmad

**Program Studi Ketatalaksanaan Pelayaran Niaga,
Program Diploma Pelayaran, Universitas Hang Tuah**

Abstrak: Pergerakan harga penutupan saham BULL cenderung mengalami variasi harga tiap hari. Investor memerlukan tindakan yang tepat, sehingga resiko yang ada dapat dikurangi dengan mengetahui naik turunnya harga saham pada masa yang akan datang dan memprediksi langkah kebijakan yang optimal untuk membuat keputusan pembelian/penjualan saham yang sesuai. Tujuan penelitian ini untuk menerapkan data mining menggunakan regresi linear untuk prediksi harga saham perusahaan pelayaran. Lokasi penelitian, yaitu di Bursa Efek Indonesia, Jakarta. Populasi dalam penelitian ini adalah semua perusahaan pelayaran yang terdaftar di Bursa Efek Indonesia. Jenis nonprobability sampling yang dipilih yaitu purposive sampling dan quota sampling. Purposive sampling yang dipakai adalah sebanyak 1 perusahaan pelayaran, yakni PT. Buana Lintas Lautan, Tbk (BULL). Quota sampling yang digunakan dalam penelitian ini adalah data time series periode harian harga pembukaan, harga tertinggi, harga terendah, harga penutupan, dan volume saham periode harian BULL selama 1 tahun 2 bulan antara bulan Juni tahun 2019 hingga bulan Juli tahun 2020. Penelitian ini menggunakan metodologi Cross Industry Standard Process for Data Mining (CRISP-DM). Proses data mining berdasarkan CRISP-DM terdiri dari 6 fase, yaitu Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment. Hasil penelitian menunjukkan masih ada selisih antara harga penutupan saham luaran data testing dengan harga penutupan saham aktual yang ada di bursa saham. Evaluasi nilai Root Mean Square Error (RMSE) menunjukkan angka plus 7,522 dari data aktual harga penutupan saham periode harian PT. BULL.

Kata kunci: data mining, regresi linear, harga saham, CRISP-DM, root mean square error

Abstract: The movement of the closing price of BULL shares tends to experience price variations every day. Investors need appropriate action, so that existing risks can be reduced by knowing the ups and downs of stock prices in the future and predicting the optimal policy steps to make appropriate share buying / selling decisions. The purpose of this study is to apply data mining using linear regression to predict the share price of shipping companies. The research location is on the Indonesia Stock Exchange, Jakarta. The population in this study are all shipping companies listed on the Indonesia Stock Exchange. The type of nonprobability sampling chosen was purposive sampling and quota sampling. The purposive sampling used is 1 shipping company, namely PT. Buana Lintas Lautan, Tbk (BULL). The sampling quota used in this study is the time series data for the daily opening price, the highest price, the lowest price, the closing price, and the volume of shares during the BULL daily period for 1 year 2 months between June 2019 and July 2020. This study uses the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. The data mining process based on CRISP-DM consists of 6 phases, namely Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The results showed that there is still a difference between the closing price of the output of the test data and the closing price of the actual shares on the stock exchange. Evaluation of the value of Root Mean Square Error (RMSE) shows the plus number 7.522 from the actual data on the closing price of shares in the daily period of PT. BULL.

Keywords: data mining, linear regression, stock price, CRISP-DM, root mean square error

Alamat korespondensi:

Ekka Pujo Ariesanto Akhmad, Program Diploma Pelayaran, Universitas Hang Tuah, Jalan A. R. Hakim 150, Surabaya. e-mail: jurnal.pdp@hangtuah.ac.id

PENDAHULUAN

Naik turunnya harga saham di pasar modal menjadi sebuah fenomena

yang menarik untuk dibicarakan berkaitan dengan isu naik turunnya nilai perusahaan itu sendiri. Fluktuasi harga

saham terjadi pada perusahaan pelayaran yang terdaftar di Bursa Efek Indonesia. Salah satu perusahaan pelayaran yang cenderung mengalami fluktuasi harga saham dalam 5 tahun terakhir adalah PT. Buana Lintas Lautan, Tbk (BULL). Pergerakan harga penutupan saham BULL cenderung mengalami variasi harga tiap hari.

Kumpulan data saham periode harian BULL membutuhkan penggalan data (data mining) untuk mencari pola atau informasi menarik dalam prediksi harga saham dengan menggunakan teknik atau metode atau algoritma tertentu.

Beberapa penelitian tentang pendekatan prediksi harga saham telah dilakukan menggunakan berbagai metode selain analisis fundamental dan analisis teknikal, antara lain Jaringan Syaraf Tiruan, Autoregressive Integrated Moving Average (ARIMA), Hidden Markov Model, Hybrid regresi dengan Algoritma Genetika, dan Support Vector Regression. Salah satu metode yang umum digunakan untuk memprediksi data adalah metode Regresi, karena memiliki keunggulan, yakni perhitungannya yang mudah.

Uraian latar belakang masalah tersebut dapat diturunkan rumusan masalah yang akan dicari jawabannya melalui penelitian ini, yaitu Bagaimana data mining menggunakan regresi linear untuk prediksi harga saham perusahaan pelayaran?

Tujuan penelitian berikut ini merupakan uraian hasil yang akan dicapai melalui penelitian, yaitu menerapkan data mining menggunakan regresi linear untuk prediksi harga saham perusahaan pelayaran.

Luaran penelitian mampu memprediksi harga saham perusahaan pelayaran periode harian dengan mengetahui perubahan nilai variabel terikat (harga penutupan), bila nilai variabel bebas (tanggal, harga pembukaan, harga tertinggi, harga

terendah, volume) dinaikkan atau diturunkan nilainya.

Penelitian yang relevan

Tabel 1 menunjukkan hasil penelitian terdahulu yang telah dilakukan peneliti-peneliti lain sebelumnya.

Persamaan penelitian yang dilakukan dengan penelitian terdahulu adalah menggunakan algoritma regresi linear.

Perbedaan penelitian yang dilakukan dengan penelitian terdahulu atau kebaruan penelitian (novelty), yaitu menggunakan perhitungan skor deviasi untuk regresi linear berganda data training dan proses data mining dengan CRISP-DM.

Cross Industry Standard Process for Data Mining (CRISP-DM)

Metodologi CRISP-DM mempunyai enam tahapan, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment (Suntoro, 2019).

Business understanding berisi tentang menentukan tujuan bisnis, menilai situasi saat ini, dan menetapkan tujuan dilakukan data mining. Data understanding merupakan kegiatan persiapan, mengevaluasi persyaratan data, dan termasuk pengumpulan data. Data preparation dilakukan setelah data dikumpulkan, data-data tersebut perlu diidentifikasi, dipilih, dibersihkan, kemudian dibangun ke dalam bentuk/format yang diinginkan. Modeling adalah aplikasi dari algoritma untuk mencari, mengidentifikasi, dan menampilkan pola. Evaluation digunakan untuk membantu pengukuran evaluasi pada model. Deployment digunakan untuk melakukan otomatisasi model atau pengembangan aplikasi, terintegrasi dengan sistem informasi manajemen atau operasional yang ada.

Tabel 1
Hasil Penelitian yang Relevan

Peneliti/ tahun	Judul	Masalah	Pendekatan/ Teori	Hasil
Padhye, Sudip and Karuna Gull (2016)	Regression Analysis for Stock Market Prediction using Weka Tool without Sentiment Analysis	prediction of stock value for upcoming week, for various companies by knowing the previous weeks stock data	<ul style="list-style-type: none"> - Linear regression - Gaussian regression - RegressionByDisc retization - RBFRegressor - MultilayerPercep tron - MLPRegressor - SMOReg - M5P (Padhye, 2016) 	<p>Technical method depends on the weekly data and the Weka forecast package. More specifically the comparison is done on the data collected for different companies using various regression methods to predict the future position of the stock market.</p> <p>This paper began by stating that irrespective of the type or the sector of the stock, predictions with less errors can be made. In our consideration of 6 stocks, KPIT belongs to IT Sector, Ashok Leyland and Bajaj Auto belongs to automotive sector, Axis Bank is a banking stock whereas LIC Housing Finance is a Finance Stock & L&T is a diversified stock company. Thus, in spite of this variation in nature of stocks, our Numerical method formula proved successful with minimal error. The comparison result of the algorithms had given more accurate results i.e. by showing less deviation from actual values collected after completion of the week under prediction. Statistical analysis drawn show the predicted results and accuracy of the methods defined.</p>
Atikah, Nabihah Hanun. (2017).	Pembuatan Aplikasi Prediksi Harga Saham Berbasis Web Menggunakan Metode Holt's dan Bayesian Belief Network: Studi Kasus di PT Bank Central Asia, Tbk.	<p>Bagaimana menentukan nilai parameter alpha dan beta yang optimal pada metode Holt's?</p> <p>Bagaimana menentukan probabilitas dari parameter-parameter yang berpengaruh terhadap harga saham pada metode Bayesian Belief Network?</p> <p>Bagaimana mengetahui aplikasi peramalan harga saham berbasis web ini dapat berfungsi dengan baik sesuai dengan deskripsi use case yang telah dibuat sebelumnya?</p>	<ul style="list-style-type: none"> - Double Exponential Smoothing Dua Parameter: Metode Holt's (Admin, 2012) - Bayesian Belief Network (Roselina, 2012)c 	<ul style="list-style-type: none"> - Dalam tugas akhir ini model peramalan dengan model <i>bayesian belief network</i> memiliki performa yang lebih baik daripada hasil peramalan menggunakan model <i>double exponential smoothing holt's</i> jika dalam hal klasifikasi arah perubahan harga saham. Hal ini dibuktikan dengan tingkat kesesuaian pola yang dihasilkan dari proses peramalan harga saham dengan menggunakan <i>bayesian belief network</i> lebih tinggi daripada <i>double exponential smoothing holt's</i> yaitu sebesar 44%. Metode <i>bayesian belief network</i> mampu mengikuti pergerakan harga saham sebesar 44%, sedangkan metode holt's sebesar 39%. - Pada tugas akhir ini, peramalan menggunakan metode holt's menghasilkan MAPE yang sangat baik yaitu sebesar 0.823%. Hal ini menunjukkan bahwa metode time series double exponential smoothing holt's memiliki performa yang baik dan cocok untuk diterapkan dalam memprediksi harga saham pada PT Bank Central Asia Tbk. Namun hasil klasifikasi menggunakan metode bayesian belief network kurang baik, yaitu hanya mampu mengikuti pergerakan harga saham sebesar 44%. Hal ini disebabkan karena data yang digunakan untuk peramalan metode bayesian belief network menggunakan data hasil interpolasi bulanan ke harian, dimana kredibilitas dari hasil interpolasi bulanan ke harian adalah rendah, sehingga menyebabkan hasil peramalan yang kurang baik.
Izzah, Abidatul dan Ratna Widyastuti. (2017)	Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression Untuk Pencegahan Data Outlier	Nilai prediksi regresi linear masih memiliki data outlier	<ul style="list-style-type: none"> - Multiple Linear Regression (Zunaidhy, d.k.k., 2008) - K-Means (Kanungo, et. al, 2002) - Moving Average (Hatidja, 2011) 	Pada penelitian ini telah dilakukan prediksi harga saham menggunakan <i>Multiple Linear Regression</i> dengan K-Means dan Moving Average. Dari hasil yang diperoleh, dapat dilihat bahwa pendekatan paling baik ditunjukkan oleh metode MLR dan MA, yakni dengan nilai MSE sebesar 15087.465, RMSE sebesar 122.831, dan MAPE sebesar 3.255

Bommareddy , Sasidhar Reddy, et. al. (2018)	Predicting The Stock Price Using Linear Regression	It is generally a dynamic market where the prices vary and it becomes difficult for an investor for predicting the prices	- Linear Regression (Bommareddy, Sasidhar Reddy, et. al., 2018)	We performed study on TCS stock prices in National Stock Exchange of India (NSEI) using Linear Regression Technique by predicting the values of Open, Close, High, and Low. Our main goal for this study is to assist stock market investors understand the future prices of TCS as predicting the stock prices is always a challenging task as the market is dynamic in nature. Future scope of this study involves considering more multiple companies from any stock exchange of different countries and also performing comparison of any different techniques of prediction so that one can understand which technique has less MSE, MAE and R2 Score.
Poornima S. P., Priyanka C. N., Reshma P., Suraj Kr Jaiswal, and Surendra Babu K N. (2019)	Stock Price Prediction using KNN and Linear Regression	Forecasting the stock prices is very challenging and complicated process because movement of price just behaves like unusual and time variation.	- K-Nearest Neighbour - Linear Regression (Poornima, 2019)	The outcome of these 2 techniques have been compared on the bases of the Confidence value. By using R2 (Coefficient of determination) we are finding the accuracy. KNN-Algorithm shows 63% of accuracy, whereas in Linear regression 98% of accuracy has been shown for daily stock prices.

Sumber: Padhye (2016), Atikah (2017), Izzah (2017), Bommareddy (2018), Poornima (2019)

Regresi Linear Berganda

Analisis regresi linear berganda sebenarnya sama dengan analisis regresi linear sederhana, hanya variabel bebasnya lebih dari satu buah. Persamaan umumnya adalah:

$$Y = a + b_1 X_1 + b_2 X_2 + \dots + b_n X_n.$$

Dengan Y adalah variabel terikat, dan X adalah variabel-variabel bebas, a adalah konstanta (intersept) dan b adalah koefisien regresi pada masing-masing variabel bebas. (konsultanstatistik, 2009)

Jika analisis regresi linear berganda dibuat, maka data setiap variabel harus tersedia. Berdasarkan data tiap variabel, peneliti dapat menemukan persamaan regresi melalui perhitungan (Sugiyono, 2012:277-292).

Uji Performa

Ukuran kesalahan pola hasil prediksi adalah kesalahan yang terjadi antara data prediksi dan data aktual. Kesalahan tersebut direpresentasikan menggunakan *mean square error* (MSE), yakni merupakan rata-rata selisih kuadrat antara nilai yang diprediksikan dengan diamati, *root mean square error*

(RMSE) merupakan akar dari MSE (Izzah, 2017). Rumus untuk MSE dan RMSE adalah sebagai berikut.

$$MSE = \frac{1}{n} \sum (Y_t - \hat{Y}_t)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum (Y_t - \hat{Y}_t)^2}$$

Sumber: Izzah (2017)

Keterangan:

n = jumlah data

Y = nilai hasil observasi

\hat{Y} = nilai hasil prediksi

t = urutan data pada database

METODE PENELITIAN

Rancangan Penelitian

Penelitian ini merupakan jenis penelitian deskriptif dengan pendekatan kuantitatif. Penelitian ini mendeskripsikan metode CRISP Data Mining menggunakan algoritma regresi linear untuk prediksi harga saham perusahaan pelayaran periode harian.

Populasi

Populasi dalam penelitian ini adalah semua perusahaan pelayaran yang terdaftar di Bursa Efek Indonesia.

Sampel

1. Purposive Sampling

Pemilihan sampel berdasarkan pertimbangan peneliti bahwa sampel yang dipilih memiliki kaya informasi. Ukuran sampel yang dipakai adalah sebanyak satu perusahaan pelayaran, yakni PT. Buana Lintas Lautan, Tbk (BULL).

2. Quota Sampling

Peneliti mengambil sampel berdasar pada pertimbangan-pertimbangan tertentu, menentukan sampel dari populasi yang mempunyai ciri-ciri tertentu sampai jumlah (kuota) yang diinginkan. Sampel yang digunakan dalam penelitian ini adalah data time series periode harian untuk harga pembukaan, harga tertinggi, harga terendah, harga penutupan, dan volume saham PT. Buana Lintas Lautan, Tbk (BULL) selama 1 tahun 2 bulan antara bulan Juni tahun 2019 hingga bulan Juli tahun 2020.

Variabel Penelitian

Penelitian ini memiliki variabel terikat, yaitu harga penutupan saham periode harian. Sedangkan variabel bebas dalam penelitian ini meliputi tanggal, harga pembukaan, harga tertinggi, harga terendah, dan volume saham. Variabel bebas diukur berdasarkan data periode harian harga saham.

Variabel penelitian selanjutnya akan dioperasionalkan. Pengukuran yang digunakan akan menghasilkan data dalam bentuk skala rasio dan nominal. Variabel tersebut disusun dengan indikator-indikator seperti pada Tabel 2.

Lokasi Penelitian

Lokasi penelitian, yaitu di Bursa Efek Indonesia, Jakarta (www.idx.co.id).

Tabel 2
Operasionalisasi variabel data mining menggunakan regresi linear untuk prediksi harga saham

Variabel	Indikator	Ukuran	Skala
Harga Penutupan [†] (Y)	Harga pada saat jam kerja bursa saham telah berakhir tiap hari	Tingkat harga	Rasio
Harga Pembukaan [†] (X1)	Harga saat awal dimulainya bursa saham tiap hari	Tingkat harga	Rasio
Harga Tertinggi [†] (X2)	Harga tertinggi yang diperoleh emiten di bursa saham/hari	Tingkat harga	Rasio
Harga Terendah [†] (X3)	Harga terendah yang diperoleh emiten di bursa saham/hari	Tingkat harga	Rasio
Volume [†] (X4)	Jumlah lembar saham yang diperdagangkan di bursa saham	Jumlah lembar saham	Rasio
Tanggal [†] (X5)	Tanggal transaksi saham	Hari/Bulan/Tahun	Nominal

Sumber: Data diolah penulis

Alur Penelitian

Penelitian ini menggunakan metodologi data mining CRISP-DM.

Gambar 1 menjelaskan enam tahapan CRISP-DM (Shearer, 2000).



Gambar 1. Tahapan CRISP-DM

Sumber: Shearer (2000)

Sumber Data

Sumber data pada penelitian ini adalah data primer seperti informasi mengenai perusahaan dan jenis jasa yang dihasilkan. Selain itu menggunakan sumber data sekunder berupa data historis harga saham tahun 2019-2020 dari Yahoo Finance dan Laporan Tahunan dari Bursa Efek Indonesia.

Teknik Pengumpulan Data

Teknik pengumpulan data yang dilakukan dalam penelitian ini adalah dokumen dan studi pustaka.

Teknik Analisis Data

Teknik analisis data penelitian ini menggunakan cara sebagai berikut.

Langkah-langkah yang akan dilakukan untuk memprediksi nilai harga saham, yaitu

1. Diawali dengan pengumpulan data dan pra-proses data.
2. Selanjutnya dilakukan proses prediksi menggunakan regresi linear berganda.
3. Proses prediksi dilakukan menggunakan *tools data mining* RapidMiner.
4. Setelah hasil prediksi diperoleh, maka hasil tersebut dianalisis dan dievaluasi. Analisis dikerjakan dengan menggambarkan grafik harga saham aktual dan grafik harga saham hasil prediksi regresi linear. Evaluasi dilakukan dengan melihat hasil Root Mean Square Error (RMSE) harga penutupan saham satu hari berikutnya hasil prediksi menurut CRISP Data Mining menggunakan algoritma regresi linear.

HASIL DAN PEMBAHASAN

Hasil Penelitian

Profil PT. Buana Listya Tama, Tbk atau PT. Buana Lintas Lautan, Tbk (BULL)

PT. Buana Lintas Lautan, Tbk menyediakan transportasi laut dan layanan terkait di Indonesia. Perusahaan beroperasi melalui Tanker Gas; Tanker Minyak, FPSO dan FSO; Tanker Kimia; dan segmen lainnya. Perusahaan bergerak dalam pengangkutan gas laut yang dicairkan, termasuk gas minyak cair, propilena, propana, dan LNG; minyak pelumas, minyak mentah, dan produk minyak bumi; dan bahan kimia cair organik dan non-organik, minyak nabati, dan lemak hewani.

Perusahaan juga menyediakan layanan fasilitas tanker apung untuk produksi, penyimpanan, dan pembongkaran produk minyak; penyewaan kapal jangka pendek, menengah, dan panjang; dan layanan

penyimpanan lepas pantai untuk sektor minyak dan gas. Selain itu, perusahaan menawarkan layanan agen pengiriman, seperti pengaturan izin dan fasilitas pelabuhan; dan penyediaan pasokan, termasuk bahan bakar, air tawar, suku cadang, layanan perbaikan, dan lainnya. Lebih lanjut, perusahaan menyediakan layanan manajemen kapal untuk kapal tanker minyak dan gas, serta untuk penyimpanan dan pembongkaran produksi terapung (Floating Production Storage and Offloading (FPSO)) / penyimpanan dan pembongkaran terapung (Floating Storage and Offloading (FSO)); dan layanan manajemen kru.

Perusahaan sebelumnya dikenal sebagai PT. Buana Listya Tama, Tbk dan berganti nama menjadi PT. Buana Lintas Lautan, Tbk pada Februari 2018. PT. Buana Lintas Lautan, Tbk didirikan pada 2005 dan berkantor pusat di Jakarta Selatan, Indonesia. PT. Buana Lintas Lautan, Tbk melakukan penawaran umum perdana (Initial Public Offering) di Bursa Efek Indonesia pada tanggal 23 Mei 2011. PT. Buana Lintas Lautan, Tbk mempunyai kode saham BULL dan diklasifikasikan ke dalam sektor infrastruktur, utilitas, dan transportasi, sub sektor transportasi.

Harga Saham BULL

Hasil penelitian meliputi perkembangan harga saham BULL selama 1 tahun 2 bulan dari tanggal 1 Juni 2019 sampai dengan 31 Juli 2020. Penelitian ini mengambil data harga saham BULL periode harian yang berasal dari halaman web finance.yahoo.com. Harga saham mengalami kenaikan sebesar Rp324,00 (harga penutupan saham tanggal 30/07/2020) – Rp191,00 (harga penutupan saham tanggal 03/06/2019) = Rp133,00 per lembar saham.

Data training untuk data mining regresi linear menggunakan harga saham BULL mulai tanggal 3 Juni 2019

hingga 29 Mei 2020. Sedangkan data testing menggunakan harga saham BULL mulai tanggal 2 Juni 2020 sampai dengan 30 Juli 2020.

Pembahasan

Penelitian ini mengerjakan tahapan CRISP-DM sebagai berikut.

Business Understanding

Investor memerlukan tindakan yang tepat untuk mengurangi resiko yang ada dengan mengetahui naik turunnya harga saham PT. BULL pada masa yang akan datang. Selain itu, investor perlu memprediksi langkah kebijakan yang sesuai untuk membuat keputusan pembelian/penjualan saham PT. BULL.

Salah satu cara yang efektif untuk memprediksi harga saham adalah dengan menggunakan teknik data mining regresi linear.

Data Understanding

Data saham PT. BULL dikumpulkan dari yahoo.finance.com. Data saham periode harian PT. BULL diperoleh penulis selama 14 bulan sejak Juni 2019 hingga Juli 2020. Data saham terdiri dari 6 atribut (5 variabel bebas dan 1 variabel terikat). Keenam atribut data saham adalah sebagai berikut.

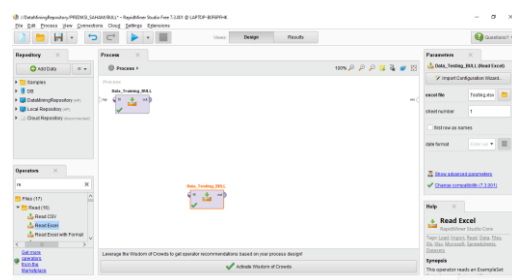
- Tanggal, berisi data tanggal transaksi saham, tipe data date_time.
- Harga Pembukaan, berisi data harga saat awal dimulainya bursa saham tiap hari, tipe data real.
- Harga Tertinggi, berisi data harga tertinggi yang diperoleh emiten di bursa saham/hari, tipe data real.
- Harga Terendah, berisi data harga terendah yang diperoleh emiten di bursa saham/hari, tipe data real.
- Volume, berisi data jumlah lembar saham yang diperdagangkan di bursa saham, tipe data real.
- Harga Penutupan, berisi data harga yang terakhir muncul pada sebuah saham sebelum bursa saham tutup, tipe data real (label).

Data Preparation

Perangkat lunak yang digunakan untuk pengolahan dan

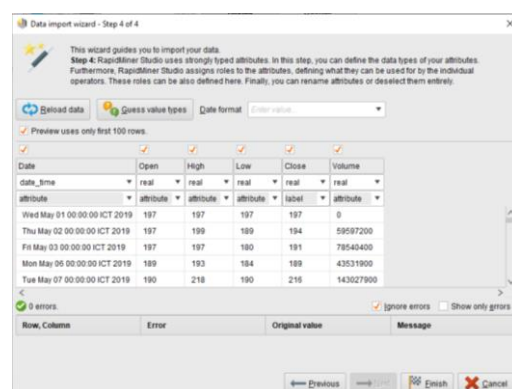
visualisasi data pada penelitian ini adalah RapidMiner 7.3. Sesuai dengan data understanding yang sudah diterangkan sebelumnya diketahui dataset harga saham PT. BULL dibagi menjadi dua kelompok yang terdiri dari 272 baris data training (data saham BULL mulai bulan Juni 2019 hingga bulan Mei 2020) dan 43 baris data testing (data saham BULL bulan Juni 2020 sampai dengan bulan Juli 2020). Persentase data training adalah 86,35% ($=272/315$) dan data testing sebesar 13,65% ($=43/315$).

Gambar 2 menjelaskan dataset harga saham PT. BULL yang telah disimpan dalam format file Excel (*.xlsx) dimasukkan ke dalam Process RapidMiner. Operator Read Excel digunakan untuk memasukkan data training saham BULL dan data testing saham BULL ke dalam Process RapidMiner.



Gambar 2. Read Excel pada RapidMiner
Sumber: file process saham BULL RapidMiner

Gambar 3 menunjukkan data saham BULL diatur tipe data dan jenis atributnya.



Gambar 3. Atur Tipe Data dan Jenis Atribut
Sumber: file process saham BULL RapidMiner

Gambar 4 menerangkan pada tab Statistics, melihat metadata dari dataset saham BULL tidak ada missing value.

Variable	Type	Missing	Statistics	Mean	StdDev	Min	Max
Close	Real	0	175	236	151.618		
Date	Date	0	May 1, 2019 12:00 AM	May 25, 2019 12:00 AM	366176 Days		
Open	Real	0	111	240	152.315		
High	Real	0	118	252	165.820		
Low	Real	0	111	224	178.676		
Volume	Real	0	0	18422400	46754106.220		

Gambar 4. Metadata dataset saham BULL
Sumber: file process saham BULL RapidMiner

Modeling

Pada tahapan modeling digunakan persamaan regresi linear berganda untuk empat prediktor atau empat variabel independen/bebas sebagai berikut.

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$

Keterangan:

Y = variabel dependen/terikat (harga penutupan (close) (Y))

X₁, X₂, X₃, X₄ = variabel independen/bebas; harga pembukaan (open) (X₁), harga tertinggi (high) (X₂), harga terendah (low) (X₃), volume (volume) (X₄)

a = konstanta (apabila nilai x sebesar 0, maka Y akan sebesar a atau konstanta)
b₁, b₂, b₃, b₄ = koefisien regresi (nilai peningkatan atau penurunan)

Berdasarkan perhitungan persamaan regresi empat prediktor untuk data training diperoleh hasil sebagai berikut.

$\Sigma X_1 = 49.590$	$\Sigma X_1^2 = 9.154.228$	$\bar{X}_1 = 182,3162$
$\Sigma X_2 = 50.543$	$\Sigma X_2^2 = 9.507.315$	$\bar{X}_2 = 185,8199$
$\Sigma X_3 = 48.600$	$\Sigma X_3^2 = 8.796.388$	$\bar{X}_3 = 178,6765$
$\Sigma X_4 = 12.717.116.900$	$\Sigma X_4^2 = 8.59973E+17$	$\bar{X}_4 = 46.754.106,25$
$\Sigma Y = 49.400$	$\Sigma Y^2 = 9.082.604$	$\bar{Y} = 181,6176$
$\Sigma X_1 Y = 9.115.385$	$\Sigma X_1 X_2 = 9.326.002$	$\Sigma X_2 X_3 = 9.141.277$
$\Sigma X_2 Y = 9.291.031$	$\Sigma X_1 X_3 = 8.971.956$	$\Sigma X_2 X_4 = 2,35431E+12$
$\Sigma X_3 Y = 8.936.174$	$\Sigma X_1 X_4 = 2,29213E+12$	$\Sigma X_3 X_4 = 2,23965E+12$
$\Sigma X_4 Y = 2,29193E+12$	$n = 272$	

Khusus variabel bebas/independen data tanggal tidak ikut dihitung dalam analisis regresi linear berganda. Alasan data tanggal tidak ikut dihitung karena tipe data tanggal adalah nominal bukan tipe data rasio. Untuk

mencari koefisien regresi b₁ (open), b₂ (high), b₃ (low), dan b₄ (volume) dapat digunakan persamaan simultan, sebagai berikut.

- $\Sigma X_1 Y = b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + b_3 \Sigma X_1 X_3 + b_4 \Sigma X_1 X_4$
- $\Sigma X_2 Y = b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2^2 + b_3 \Sigma X_2 X_3 + b_4 \Sigma X_2 X_4$
- $\Sigma X_3 Y = b_1 \Sigma X_1 X_3 + b_2 \Sigma X_2 X_3 + b_3 \Sigma X_3^2 + b_4 \Sigma X_3 X_4$
- $\Sigma X_4 Y = b_1 \Sigma X_1 X_4 + b_2 \Sigma X_2 X_4 + b_3 \Sigma X_3 X_4 + b_4 \Sigma X_4^2$

Dengan metode skor deviasi diperoleh hasil sebagai berikut.

$$\Sigma \Sigma X_1^2 = 9.154.228 - (49.590)^2 / 272 = 113.169$$

$$\Sigma \Sigma X_2^2 = 9.507.315 - (50.543)^2 / 272 = 115.422$$

$$\Sigma \Sigma X_3^2 = 8.796.388 - (48.600)^2 / 272 = 112.711$$

$$\Sigma \Sigma X_4^2 = 8,59973E+17 - (12.717.116.900)^2 / 272 = 2,65396E+17$$

$$\Sigma \Sigma Y^2 = 9.082.604 - (49.400)^2 / 272 = 110.692$$

$$\Sigma \Sigma X_1 Y = 9.115.385 - (49.590)(49.400) / 272 = 108.966$$

$$\Sigma \Sigma X_2 Y = 9.291.031 - (50.543)(49.400) / 272 = 111.530$$

$$\Sigma \Sigma X_3 Y = 8.936.174 - (48.600)(49.400) / 272 = 109.556$$

$$\Sigma \Sigma X_4 Y = 2,29193E+12 - (12.717.116.900)(49.400) / 272 = -17.722.848.750$$

$$\Sigma \Sigma X_1 X_2 = 9.326.002 - (49.590)(50.543) / 272 = 111.195$$

$$\Sigma \Sigma X_1 X_3 = 8.971.956 - (49.590)(48.600) / 272 = 111.390$$

$$\Sigma \Sigma X_1 X_4 = 2,29213E+12 - (49.590)(12.717.116.900) / 272 = -26.406.128.938$$

$$\Sigma \Sigma X_2 X_3 = 9.141.277 - (50.543)(48.600) / 272 = 110.432$$

$$\Sigma \Sigma X_2 X_4 = 2,35431E+12 - (50.543)(12.717.116.900) / 272 = -8.782.792.194$$

$$\Sigma \Sigma X_3 X_4 = 2,23965E+12 - (48.600)(12.717.116.900) / 272 = -32.599.563.750$$

Hasil skor deviasi dimasukkan ke dalam persamaan simultan sebagai berikut.

$$1) 108.966 = 113.169 b_1 + 111.195 b_2 + 111.390 b_3 + -26.406.128.938 b_4$$

$$2) 111.530 = 111.195 b_1 + 115.422 b_2 + 110.432 b_3 + -8.782.792.194 b_4$$

$$3) 109.556 = 111.390 b_1 + 110.432 b_2 + 112.711 b_3 + -32.599.563.750 b_4$$

$$4) -17.722.848.750 = -26.406.128.938 b_1 + -8.782.792.194 b_2 + -32.599.563.750 b_3 + 2,65396E+17 b_4$$

$$(1a) -4,12654E-06 = -4,28571E-06 b_1 + -4,21095E-06 b_2 + -4,21834E-06 b_3 + b_4$$

$$(2a) -1,26987E-05 = -1,26606E-05 b_1 + -1,31418E-05 b_2 + -1,25737E-05 b_3 + b_4$$

$$(3a) -3,36066E-06 = -3,41692E-06 b_1 + -3,38753E-06 b_2 + -3,45744E-06 b_3 + b_4$$

$$(4a) -6,67789E-08 = -9,94971E-08 b_1 + -3,30932E-08 b_2 + -1,22834E-07 b_3 + b_4$$

$$(5)(1a - 2a) 8,57215E-06 = 8,37484E-06 b_1 + 8,93088E-06 b_2 + 8,35534E-06 b_3$$

$$(6)(2a - 3a) -9,33804E-06 = -9,24364E-06 b_1 + -9,7543E-06 b_2 + -9,11624E-06 b_3$$

$$(7)(3a - 4a) -3,29388E-06 = -3,31742E-06 b_1 + -3,35444E-06 b_2 + -3,33461E-06 b_3$$

$$(5a) 1,03 = 1,00233428 b_1 + 1,068883033 b_2 + b_3$$

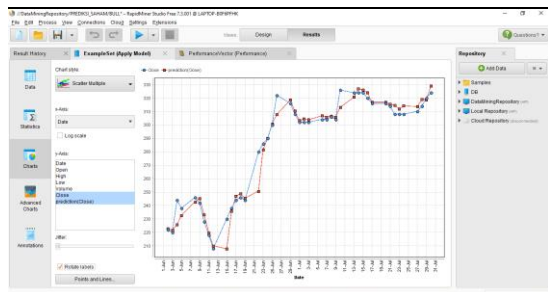
$$(6a) 1,02 = 1,013974671 b_1 + 1,069992171 b_2 + b_3$$

$$(7a) 0,987786964 = 0,994846336 b_1 + 1,005947209 b_2 + b_3$$

$$(8) (5a - 6a) 1,62E-03 = -1,16E-02 b_1 + -1,11E-03 b_2$$

$$(9) (6a - 7a) 3,65E-02 = 1,91E-02 b_1 + 6,40E-02 b_2$$

model menghasilkan grafik garis regresi linear berganda untuk data training dan data testing pada Gambar 6. Sedangkan hubungan dengan operator performance akan menghasilkan nilai RMSE.



Gambar 6. Grafik garis data testing terhadap data training harga penutupan saham BULL
Sumber: file process saham BULL RapidMiner

Gambar 6 menunjukkan grafik garis dari persamaan regresi linear berganda untuk data training dan data testing harga saham penutupan PT. BULL. Grafik garis mempunyai keterangan sumbu x untuk tanggal periode harian dan sumbu y untuk harga penutupan saham periode harian. Periode harian menggunakan data time series dari tanggal 2 Juni 2020 hingga 30 Juli 2020.

Evaluation

Uji coba prediksi variabel terikat yakni, harga penutupan saham (Y) PT. BULL periode harian tanggal 30 Juli 2020 menggunakan data variabel bebas tanggal 30 Juli 2020 sebagai berikut.

Harga pembukaan (X_1) = 322;
Harga tertinggi (X_2) = 334;
Harga terendah (X_3) = 322;
Volume (X_4) = 155.285.400

Kemudian data variabel bebas dimasukkan ke persamaan regresi linear berganda dari data testing sebagai berikut.

$$\begin{aligned} Y &= 11,622 + -0,351 X_1 + 0,730 X_2 + 0,594 X_3 + -5,258E-8 X_4 \\ &= 11,622 + -0,351.(322) + 0,730.(334) \\ &\quad + 0,594.(322) + -5,258E-8.(155285400) \\ &= 325,5231 \end{aligned}$$

Jadi harga saham PT. BULL keluaran persamaan regresi linear berganda data testing untuk tanggal 30 Juli 2020 adalah 325,5231 rupiah.

Namun, data aktual harga penutupan saham PT. BULL pada tanggal 30 Juli 2020 adalah 324 rupiah, sehingga prediksi harga penutupan saham PT. BULL memiliki selisih 1,5231 rupiah antara data prediksi tanggal 30 Juli 2020 dengan data aktual tanggal 30 Juli 2020.

Evaluasi mengukur kesalahan yang terjadi antara data prediksi dan data aktual direpresentasikan menggunakan *root mean square error* (RMSE), yaitu akar rata-rata selisih kuadrat antara nilai yang diprediksikan dengan diamati.

Luaran nilai RMSE dari data mining regresi linear data testing dengan Rapidminer untuk harga penutupan saham PT. BULL sama dengan plus 7,522 dari data aktual harga penutupan saham PT. BULL di bursa saham.

Deployment

Penyebaran (deployment) diterapkan kepada investor yang investasi saham PT. BULL. Namun, investor belum bisa menggunakan model regresi linear berganda data testing luaran penelitian ini untuk pembelian atau penjualan saham PT. BULL.

Ada beberapa hal yang perlu diperhatikan pengembangan data mining sebagai berikut.

- Data time series periode harian atau dataset bisa diubah jumlah datanya menjadi lebih besar, misal 3 atau 5 tahun terakhir.
- Tahap modelling data mining bisa menggunakan algoritma regresi linear dengan pengembangan atau algoritma lain untuk meminimalkan nilai RMSE antara data prediksi dan data aktual harga saham.

KESIMPULAN DAN SARAN

Kesimpulan

Proses data mining menggunakan Cross Industry Standard Process for Data Mining (CRISP-DM). Data mining menggunakan data time series periode harian mulai bulan Juni tahun 2019 hingga bulan Juli tahun 2020. Data mining telah menghasilkan persamaan regresi linear berganda dan grafik garis regresi linear berganda untuk data training dan data testing. Uji coba prediksi harga penutupan saham periode harian perusahaan pelayaran dilakukan dengan menggunakan persamaan regresi linear berganda untuk data testing. Hasil uji coba menunjukkan masih ada selisih antara harga penutupan saham luaran data testing dengan harga penutupan saham aktual yang ada di bursa saham. Evaluasi nilai Root Mean Square Error (RMSE) menunjukkan angka plus 7,522 dari data aktual harga penutupan saham periode harian PT. BULL.

Saran

Data mining dapat menggunakan algoritma selain regresi linear, misal jaringan syaraf tiruan untuk meminimalkan selisih antara harga penutupan saham periode harian perusahaan pelayaran luaran prediksi data testing dengan harga penutupan saham periode harian yang sebenarnya.

DAFTAR PUSTAKA

- Admin. Teknik Peramalan S1 Pertemuan 5. (2012). Diakses dari http://ocw.stikom.edu/course/download/2012/11/16-03-2012.15.12.32_745_410103096_Teknik-Peramalan-S1-SI_P1_Pert5_1.doc.
- Arikunto, Suharsimi. (2010). *Prosedur Penelitian : Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Atikah, Nabihah Hanun. (2017). Pembuatan Aplikasi Prediksi Harga Saham Berbasis Web Menggunakan Metode Holt's dan Bayesian Belief Network: Studi Kasus di PT Bank Central Asia, Tbk. Tugas Akhir. Departemen Sistem Informasi, Fakultas Teknologi Informasi, Institut Sepuluh Nopember Surabaya.
- B. Heizer and Render. (2014). *Operation Management Sustainability and Suply Chain Management*, 11th Edition, Pearson.
- Bommareddy, Sasidhar Reddy, et. al. (2018). Predicting The Stock Price Using Linear Regression. *International Journal of Advanced Research in Computer Science*. Volume 9, Special Issue No. 3, May 2018, 81-85.
- D. Hatidja. (2011). Penerapan Model Arima Untuk Memprediksi Harga Saham PT Telkom Tbk. *Jurnal Ilmiah Sains*, Vol. 11, No. 1, Pp. 116–123.
- Duwi Consultant. (2011). Analisis Regresi Linear Sederhana. Diakses dari <http://duwiconsultant.blogspot.com/2011/11/analisis-regresi-linier-sederhana.html> tanggal 5/12/2019.
- Izzah, Abidatul dan Ratna Widyastuti. (2017). Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression Untuk Pencegahan Data Outlier. *KINETIK*, Vol. 2, No. 3, Agustus 2017, Hal. 141-150.
- Konsultanstatistik. (2009). Regresi Linear Berganda. Diakses dari <http://www.konsultanstatistik.com/2009/03/regresi-linear.html> tanggal 5/12/2019.
- Padhye, Sudip and Karuna Gull. (2016). Regression Analysis for Stock Market Prediction using Weka Tool without Sentiment Analysis. *Sixth International Conference on Computational Intelligence and Information Technology – CIIT 2016*

- Poornima S. P., Priyanka C. N., Reshma P., Suraj Kr Jaiswal, and SurendraBabu K. N. (2019). Stock Price Prediction using KNN and Linear Regression. *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249–8958, Volume-8, Issue-5S, May, 2019.
- R. Zunaidhi, W. Saputra, N. Sari. (2008). Aplikasi Peramalan Penjualan Menggunakan Metode Regresi Linier. *Jurnal Scan*, Vol. 2, No. 3, pp. 41-45.
- Roselina. (2012). Aplikasi Diagnosa Penyakit Asma Menggunakan Bayesian Network Berbasis Web. *Teknik Informatika*. Vol. 1, September 2012, Riau.
- Shearer, Colin. (2000). The CRIPS-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, Vol. 5, No. 4 Fall 2000.
- Sugiyono. (2012). *Metode Penelitian Bisnis*. Bandung: Penerbit ALFABETA.
- Suntoro, Joko. (2019). Data Mining Algoritma dan Implementasi dengan Pemrograman PHP. Jakarta: PT. Elex Media Komputindo.
- T. Kanungo, D. Mount, N. Netanyahu. (2002). An Efficient K-Means Clustering Algorithm: Analysis and Implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7.
- Tim Riset CNBC Indonesia. (2019). di era tol laut emiten kapal berpeluang cetak keajaiban. Diakses dari <https://www.cnbcindonesia.com/market/20190225205153-17-57543/di-era-tol-laut-emiten-kapal-berpeluang-cetak-keajaiban/2> tanggal 24/12/2019