



Analisis Pola Prediksi Data *Time Series* menggunakan *Support Vector Regression*, *Multilayer Perceptron*, dan Regresi Linear Sederhana

Ika Oktavianti¹, Ermatita², Dian Palupi Rini²

Magister Teknik Informatika, Ilmu Komputer, Universitas Sriwijaya

¹ioktavianti@vittal.org, ²ermatitaz@yahoo.com, ³dprini@unsri.ac.id

Abstract

Licensing services is one of the forms of public services that important in supporting increased investment in Indonesia and is currently carried out by the Investment and Licensing Services Department. The problems that occur in general are the length of time to process licenses and one of the contributing factors is the limited number of licensing officers. Licensing data is a time series data which have monthly observation. The Artificial Neural Network (ANN) and Support Vector Machine (SVR) is used as machine learning techniques to predict licensing pattern based on time series data. Of the data used dataset 1 and dataset 2, the sharing of training data and testing data is equal to 70% and 30% with consideration that training data must be more than testing data. The result of the study showed for Dataset 1, the ANN-Multilayer Perceptron have a better performance than Support Vector Regression (SVR) with MSE, MAE and RMSE values is 251.09, 11.45, and 15.84. Then for dataset 2, SVR-Linear has better performance than MLP with values of MSE, MAE and RMSE of 1839.93, 32.80, and 42.89. The dataset used to predict the number of permissions is dataset 2. The study also used the Simple Linear Regression (SLR) method to see the causal relationship between the number of licenses issued and licensing service officers. The result is that the relationship between the number of licenses issued and the number of service officers is less significant because there are other factors that affect the number of licenses.

Keywords: *pattern prediction, time series data, multilayer perceptron, support vector regression, simple linear regression*

Abstrak

Layanan perizinan adalah salah satu bentuk layanan publik yang penting dalam mendukung peningkatan investasi di Indonesia dan saat ini dilakukan oleh Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu. Masalah yang terjadi secara umum adalah lamanya waktu untuk memproses izin dan salah satu faktor penyebab adalah terbatasnya jumlah petugas perizinan. Data perizinan adalah data deret waktu yang diobservasi setiap bulan. *Artificial Neural Network* (ANN) dan *Support Vector Regression* (SVR) digunakan sebagai teknik pembelajaran mesin untuk memprediksi pola perizinan berdasarkan data deret waktu. Dari data *time series* yang digunakan Dataset 1 dan Dataset 2, pembagian data pelatihan dan data pengujian sama dengan 70% dan 30%, dengan pertimbangan bahwa data *training* harus lebih banyak dibandingkan data *testing*. Hasil penelitian menunjukkan bahwa untuk Dataset 1, ANN-Multilayer Perceptron memiliki kinerja yang lebih baik daripada SVR dengan nilai MSE, MAE dan RMSE adalah 251.09, 11.45, dan 15.84. Kemudian untuk Dataset 2, SVR-Linear memiliki kinerja lebih baik dibandingkan MLP dengan nilai MSE, MAE dan RMSE sebesar 1839.93, 32.80, dan 42.89. Dataset yang digunakan untuk prediksi jumlah perizinan adalah dataset 2. Penelitian juga menggunakan metode *Simple Linear Regression* (SLR) untuk melihat hubungan sebab akibat antara jumlah izin yang terbit dengan petugas pelayanan perizinan. Hasilnya hubungan jumlah izin yang terbit dengan jumlah petugas pelayanan kurang signifikan karena terdapat faktor lain yang mempengaruhi jumlah izin.

Kata kunci: pola prediksi, data *time series*, *multilayer perceptron*, *support vector regression*, regresi linear sederhana

1. Pendahuluan

Layanan perizinan adalah salah satu bentuk layanan publik yang penting dalam meningkatkan investasi dan pembangunan di suatu daerah. Reformasi birokrasi perizinan telah diimplementasikan sejak lama sebagai bentuk upaya pemerintahan dalam memberikan pelayanan prima kepada masyarakat. Hal ini dikarenakan berbagai masalah dalam perizinan, seperti waktu pemrosesan yang lambat, persyaratan yang tumpang tindih, pungutan liar terhadap investor dan lain-lain.

Peraturan Presiden Nomor 97 Tahun 2014 tentang Pelayanan Terpadu Satu Pintu di Bidang Penanaman Modal dikeluarkan untuk memberikan layanan yang lebih baik kepada warga negara serta mempersingkat proses layanan agar lebih cepat, mudah, murah, transparan, pasti dan terjangkau yang diimplementasikan dalam Pelayanan Terpadu Satu Pintu yang selanjutnya disingkat menjadi PTSP. PTSP adalah layanan terintegrasi mulai dari tahap permohonan hingga penyelesaian produk perizinan. Menindaklanjuti Peraturan Presiden di atas, ditetapkan bahwa PTSP berada dibawah ruang lingkup Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu dan seluruh provinsi di Indonesia memiliki nomenklatur yang sama, yaitu DPMPTSP. DPMPTSP Provinsi Sumatera Selatan saat ini melayani 140 jenis perizinan dan non perizinan.

Berdasarkan masalah perizinan yang telah dijelaskan, penelitian dilakukan untuk mengidentifikasi pola prediksi data *time series* perizinan. Studi ini mengidentifikasi data perizinan yang diterbitkan secara berkala setiap bulan. Identifikasi berguna untuk menemukan pengetahuan dalam data. Berbagai pendekatan untuk prediksi digunakan dalam penelitian ini yaitu teknik *machine learning*. *Support Vector Regression* (SVR) and *Neural Network* adalah salah satu metode yang efektif dan menghasilkan algoritma terbaik dengan data yang terbatas [1].

Beberapa penelitian telah membuktikan bahwa metode SVR dan ANN mampu melakukan peramalan. Salah satunya adalah penelitian tentang prognosis kerusakan bantalan gelinding dengan menggunakan metode *Support Vector Regression* (SVR). Penelitian ini dilakukan dengan menggunakan objek bantalan gelinding karena bantalan gelinding merupakan komponen yang mampu membuat sebuah mesin terus berputar atau bekerja. Dengan melakukan prognosis terhadap kerusakan bantalan gelinding dapat mengoptimalkan biaya perawatan mesin karena bisa mengetahui sisa umur fungsi bantalan gelinding sebelum bantalan gelinding tersebut rusak [2].

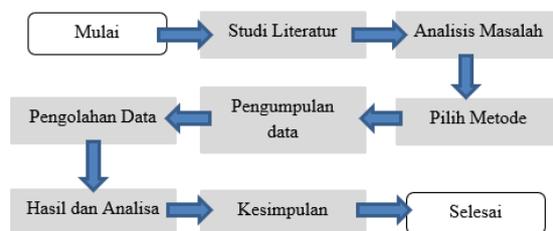
Penelitian lainnya adalah penelitian mereka bahwa model *Support Vector Regression* dapat memprediksi

lebih akurat daripada model *Moving Average* (MA). Model mereka dapat menghasilkan hampir 90% prediksi akurat untuk perkiraan curah hujan [3].

Salah satu karakteristik utama SVR untuk mencapai kinerja umum adalah upaya untuk meminimalkan batasan kesalahan umum dan untuk menghindari masalah kelebihan beban. Kesalahan umum adalah perbedaan antara risiko empiris pada data pelatihan dan risiko hasil hipotesis. Dengan kata lain, mengukur seberapa akurat model untuk memprediksi nilai hasil dari data sebelumnya [4]. Metode *Simple Linear Regression* (SLR) juga digunakan dalam penelitian ini untuk membuktikan apakah ada hubungan yang signifikan antara jumlah izin yang terbit dengan jumlah petugas pelayanan. Diharapkan hasil penelitian ini akan membantu Pemerintah Provinsi Sumatera Selatan dalam merencanakan pengambilan keputusan dan kebijakan terkait peningkatan pelayanan dan Indeks Kepuasan Masyarakat (IKM).

2. Metode Penelitian

Metode penelitian merupakan suatu proses tata cara berurutan yang dilakukan untuk kepentingan penelitian. Adapun metode yang digunakan untuk membantu dalam penelitian yang berjudul Prediksi Pola Perizinan menggunakan data *time series* dan dengan menggunakan *Support Vector Regression* (SVR) dan *Artificial Neural Network* (ANN) - *Multilayer Perceptron* pada Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu Provinsi Sumatera Selatan. Proses yang dilakukan pada penelitian dapat dilihat pada alur gambar 1.

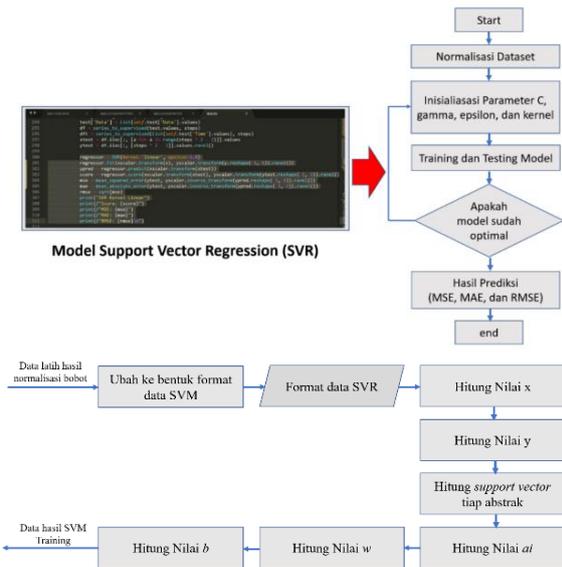


Gambar 1. Alur Proses Penelitian

2.1. Pemodelan SVR

Penelitian ini menggunakan dataset perizinan yang telah diterbitkan setiap bulan dari Agustus 2009 hingga Januari 2019. Data dibagi menjadi *training set* dan *testing set* secara default. 70% dari data digunakan sebagai *training set* dan 30% dari data digunakan sebagai *testing set*. *Training data* akan digunakan sebagai input untuk *Multilayer Perceptron* (MLP) dan *Support Vector Regression* (SVR), sedangkan *testing data* akan digunakan untuk prediksi. Kinerja MLP-ANN dan SVR diukur dengan *Mean Square Error* (MSE), *Mean Absolute Error* (MAE) dan *Root Mean Square Error* (RMSE). Gambar 2 menunjukkan hasil

coding untuk menemukan hasil prediksi melalui Bahasa pemrograman *python*. Setelah model optimal, maka didapat hasil MAE, MSE, RMSE untuk menentukan apakah tingkat kesalahan yang terjadi dengan dataset tinggi atau rendah. Kemudian akurasi MLP-ANN dan SVR akan dibandingkan untuk melihat model yang lebih cocok untuk data *time series* dari dataset perizinan. Pada proses pelatihan SVR bertujuan untuk menemukan vektor α , nilai w dan konstanta b untuk mendapatkan *hyperplane* terbaik. Dalam proses pelatihan dibutuhkan satu set data *training* dan data *testing*.

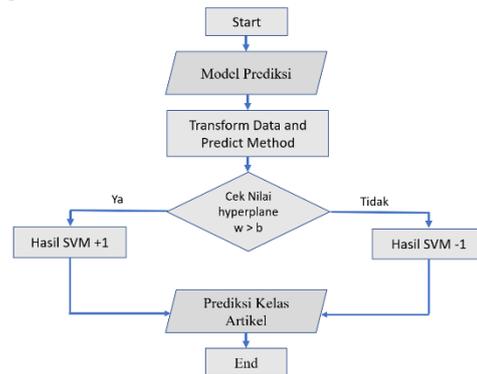


Gambar 2. Model Support Vector Regression

Data hasil pembobotan dan normalisasi diubah ke dalam format data SVR. Masukan yang pertama pada *kelas* adalah +1 atau -1 menyatakan dua label awalan yang diberikan. Tahap selanjutnya menghitung nilai x . nilai X pada tabel akan digunakan untuk perhitungan kernel. Selanjutnya yaitu melakukan kernelisasi menggunakan fungsi Kernel linier. Kemudian tahap selanjutnya adalah melakukan perhitungan terhadap y . Untuk nilai y adalah nilai dari label yang diberikan. Tahap selanjutnya mengubah setiap abstrak menjadi nilai vektor (*support vector*) = $\frac{x}{y}$ agar mendapatkan nilai ai . Setelah didapatkan nilai ai , maka akan didapatkan nilai w dan b . Model Support Vector Regression dapat dilihat pada gambar 6. Dalam Support Vector Regression, input x pertama-tama dipetakan ke dalam ruang fitur menggunakan model linear untuk membangun input dalam ruang dimensi tinggi [5].

Setelah mendapatkan nilai w dan b atau *hyperplane* dari hasil SVM Training, selanjutnya dapat menentukan data abstrak masuk dalam kelas positif atau negatif dengan nilai w dan *hyperplane* tersebut. Jika nilai hasil uji lebih besar dari nilai *hyperplane* maka abstrak tersebut masuk dalam kelas positif, jika lebih kecil dari

nilai *hyperplane* maka abstrak tersebut masuk dalam kelas negatif. Flowchart SVR Testing dapat dilihat pada gambar 3.



Gambar 3. Flowchart Support Vector Regression

2.2. Pemodelan MLP

ANN adalah model komputasi atau matematika yang terinspirasi oleh proses otak manusia [6]. Sistem dirancang untuk memecahkan masalah yang lebih kompleks. *Neural Network* terdiri dari beberapa neuron, neuron ini akan mengubah informasi yang diterima melalui outletnya ke neuron lain. Dalam *Neural Network*, hubungan ini dikenal sebagai bobot. Informasi ini disimpan pada nilai tertentu pada berat itu. Bobot masing-masing koneksi bersifat dinamis sehingga input yang diberikan menghasilkan output yang diinginkan. Model *Artificial Neuron* terdiri dari fungsi aktivasi. Fungsi aktivasi yang paling sering digunakan adalah fungsi *hyperbolic tangent* dan *sigmoidal functions*. *Feed forward neural networks* adalah yang terkenal di berbagai aplikasi di antara berbagai jenis koneksi untuk neuron buatan. Ia juga dikenal sebagai *Multilayer Perceptron (MLP)*.

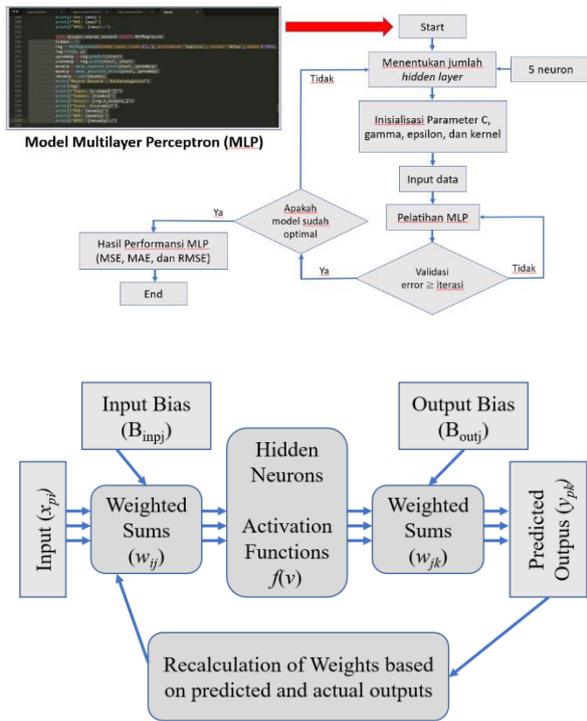
Multilayer Perceptron (MLP) adalah *feed forward neural network* yang terdiri dari beberapa neuron yang terhubung dengan menghubungkan bobot. Setiap bobot terhubung ke setiap bobot pada lapisan berikutnya. Setiap MLP terdiri dari minimal tiga lapis dari input layer. Neuron-neuron ini tersusun dalam lapisan-lapisan yang terdiri dari satu *input layer*, satu atau lebih *hidden layer*, dan satu *output layer*. *Input layer* menerima sinyal dari luar, kemudian meneruskannya ke *hidden layer* pertama, yang akan berlanjut hingga mencapai *output layer*.

Tahapan pemodelan *Multilayer Perceptron Neural Network* dimulai dengan penentuan jumlah *hidden layer*. Tahapan pemodelan dapat dilihat pada gambar 4.

2.3. Pemodelan SLR

Model persamaan Regresi Linear Sederhana dijelaskan dalam rumus 1.

$$y = a + b x \quad (1)$$



Gambar 4. Model Multilayer Perceptron

Untuk mendapatkan hasil persamaan regresi tersebut, hal pertama yang dilakukan adalah menentukan *dataset* sampel yang akan digunakan dan peneliti akan menggunakan jumlah izin dataset 2 yaitu dari bulan Februari 2016 sampai dengan bulan Januari 2019 yaitu sebanyak 36 data.

x adalah Jumlah Izin dan y adalah Jumlah Pegawai Pelayanan Perizinan (SDM). Untuk mengetahui nilai persamaannya, maka dicari nilai a dan b . Setelah dilakukan perhitungan, didapat nilai a dan b sebagai berikut:

$$y = 13,951 + 0,010832x \quad (2)$$

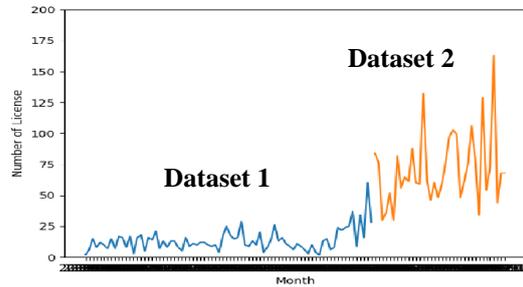
3. Hasil Analisis dan Pembahasan

3.1. Plot Data

Dalam penelitian ini, plot data dilakukan untuk menentukan *trend* pengurusan perizinan yang menjadi kewenangan provinsi setiap bulannya. Banyaknya perizinan yang diselesaikan setiap bulannya diharapkan dapat menggambarkan dan membantu menghasilkan model, variasi *trend* musiman bagi perusahaan untuk mengurus perizinannya. Plot data juga digunakan untuk menentukan teknik yang tepat dalam melakukan peramalan. Plot data *time series* dari dataset perizinan yang terbit setiap bulan dapat dilihat pada gambar 5.

Gambar 5 diatas menunjukkan bahwa plot data *time series* yang digunakan dalam penelitian ini adalah plot

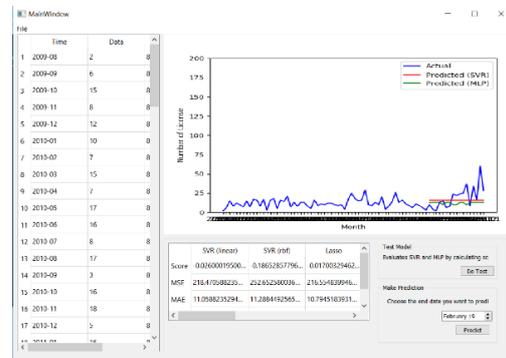
data *time series* dengan 2 (dua) dataset berulang atau musiman setiap bulan. Pola ini membentuk siklus yang dapat membantu melakukan peramalan (*forecasting*).



Gambar 5. Plot Data *Time Series*

3.2. Hasil Analisis SVR dan MLP

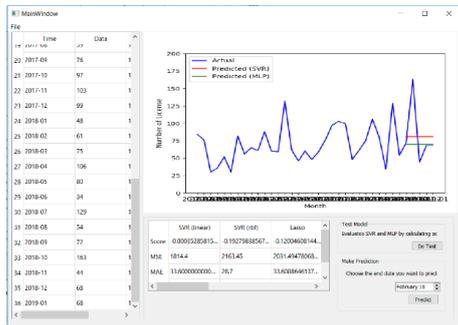
Dataset 1 terdiri dari 78 *actual data* jenis izin, data yang dilatih adalah 70% yaitu sebanyak 55 data dan data yang diuji untuk menghitung MAE, MSE, dan RMSE hanya $\pm 30\%$ yaitu sebanyak 23 data. Hasil analisis dataset 1 adalah tingkat akurasi MLP lebih baik dibandingkan SVR-Linear yaitu *score* -0.17 dan nilai MSE, MAE dan RMSE sebesar 251.09, 11.45, dan 15.84. Terbukti nilai *error* MLP lebih rendah dibandingkan dengan SVR-Linear yaitu sebesar 15.84%. Namun apabila dibandingkan dengan SVR fungsi kernel lainnya, maka SVR-Ridge lebih akurat dibandingkan dengan MLP dengan *score* -0.011 dan nilai MSE, MAE dan RMSE sebesar 215.33, 10.76, 14.67. Nilai *error* SVR-Ridge lebih kecil dibandingkan MLP yaitu 14.67%. Grafik hasil perhitungan SVR dan MLP untuk dataset 1 dapat dilihat pada gambar 6.



Gambar 6. Hasil SVR dan MLP Dataset 1 menggunakan Bahasa Program Python

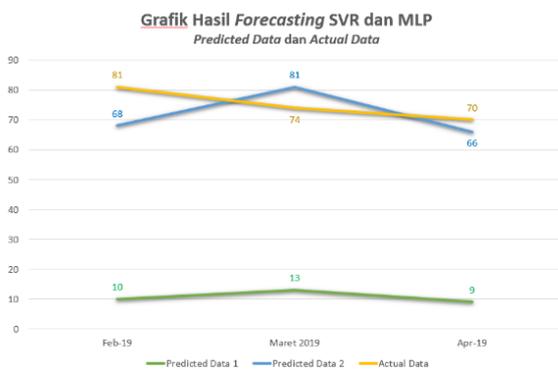
Dataset 2 terdiri dari 36 *actual data* jenis izin yang ada, data yang dilatih adalah 70% yaitu sebanyak 25 data dan data yang diuji untuk menghitung MAE, MSE, dan RMSE hanya $\pm 30\%$ yaitu sebanyak 11 data. Hasil analisis dataset 2 adalah tingkat akurasi SVR-Linear lebih baik dibandingkan MLP yaitu *score* -0.014 dan nilai MSE, MAE dan RMSE sebesar 1839.93, 32.80, dan 42.89 dan nilai SVR fungsi kernel lainnya. Terbukti nilai *error* SVR-linear lebih rendah

dibandingkan dengan MLP SVR fungsi kernel lainnya yaitu sebesar 42.89%. Hasil prediksi dan hasil tes data dapat dilihat pada gambar 7.



Gambar 7. Hasil SVR dan MLP Dataset 2 menggunakan Bahasa Program Python

Hasil prediksi (*Predicted*) izin yang akan terbit selama 3 bulan berikutnya apabila dibandingkan dengan hasil yang sebenarnya (*actual*) dapat dilihat pada grafik gambar 8.



Gambar 8. Grafik Forecasting SVR dan MLP

Perhitungan yang dilakukan dengan menggunakan kedua *dataset time series* menghasilkan prediksi jumlah perizinan yang akan terbit di bulan dan tahun mendatang sangat berbeda. Hal ini disebabkan oleh adanya perubahan musim yang mengakibatkan hasil prediksi menggunakan *dataset 1* tidak sesuai lagi dengan perkembangan perizinan sekarang. Sehingga walaupun nilai *error* berdasarkan *dataset 1* lebih kecil dibandingkan *dataset 2*, tetapi yang bisa dijadikan acuan dalam memprediksi perizinan tetap berdasarkan hasil perhitungan *dataset 2*. Terbukti pada saat dibandingkan dengan hasil sebenarnya, prediksi menggunakan *dataset 2* tidak terlalu jauh dengan hasil yang sebenarnya.

3.3. Hasil Forecasting menggunakan SLR

Setelah data dimasukkan ke dalam persamaan (10) dengan menggunakan data random dari seluruh data *testing* sebanyak 36 data, maka hasil peramalan yang didapat bahwa berdasarkan jumlah izin yang terbit rata-rata sebanyak 71 izin diperlukan petugas pelayanan (SDM) sebanyak 15 orang atau 1 izin harus dikerjakan

oleh 4 orang agar pelayanan bisa optimal dan prima. Hal ini diperlukan agar kinerja proses pelayanan perizinan dapat berjalan maksimal. Tabel 1 menunjukkan hasil perhitungan dan analisa apakah ada hubungan yang signifikan antara jumlah izin yang terbit dan jumlah petugas pelayanan perizinan.

Tabel 1. Hasil Regresi menggunakan SPSS

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,258 ^a	,066	0.39	1,209

a. Predictors: (Constant), izin

ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	3,531	1	3,531	2,416	,129 ^a
Residual	46,692	34	1,462		
Total	53,222	35			

a. Dependent Variable: Petugas
b. Predictors: (Constant), Izin

Coefficients ^a					
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	13,951	,536		26,052	,000
Izin	,011	,007	,258	1,554	,129

a. Dependent Variable: Petugas

Hasil analisis berdasarkan perhitungan Tabel 1 menggunakan program aplikasi *Statistical Product and Service Solutions* (SPSS), antara lain:

1. Persamaan yang dihasilkan dari perhitungan secara manual regresi linear sederhana sudah benar dan sesuai, yaitu a sebesar 13,951 dan b sebesar 0,011.
2. Nilai R Square 0,066 menunjukkan bahwa hubungan antara jumlah izin yang terbit dengan jumlah petugas yang melayani perizinan kurang signifikan karena hanya sebesar 6,6%. Ada faktor lain yang lebih mempengaruhi jumlah izin yang terbit (x), tidak hanya faktor jumlah petugas pelayanan (y).
3. Hasil *standard error of estimate* yaitu sebesar 1,209 menunjukkan bahwa tingkat *error* untuk melakukan peramalan adalah sebesar 12,09% yang berarti cukup tinggi. Hal ini dikarenakan standar *error* dalam suatu peramalan menggunakan *simple linear regression* adalah sebesar 5%.

4. Kesimpulan

Penelitian ini menggunakan data *time series* dengan metode *Support Vector Regression* (SVR) dan *Multilayer Perceptron* (MLP) untuk memprediksi pola perizinan. Hasil penelitian menunjukkan bahwa:

1. Prediksi menggunakan data *time series* yang cukup banyak membuktikan bahwa metode MLP lebih

- akurat dibandingkan dengan metode SVR, dimana nilai akurasinya adalah sebesar -0.17 dan nilai *Mean Absolute Error (MAE)* dan *Root Mean Square Error (RMSE)* adalah 11,45 dan 15,84.
2. Prediksi menggunakan data *time series* yang sedikit membuktikan bahwa metode SVR lebih akurat dibandingkan dengan metode MLP, dimana nilai akurasinya adalah sebesar -0.014 dan nilai *Mean Absolute Error (MAE)* dan *Root Mean Square Error (RMSE)* adalah 32,80 dan 42,89.
 3. Metode SVR dan MLP-ANN terbukti cukup baik digunakan untuk meramalkan suatu hal dan dengan dukungan data *time series* yang cukup banyak metode ini dapat menghasilkan pola prediksi yang lebih akurat.
 4. Hasil peramalan menggunakan *Simple Linear Regression (SLR)*, menunjukkan bahwa jumlah petugas pelayanan masih kurang dan pemerintah perlu menambah petugas pelayanan agar kualitas pelayanan perizinan dan Indeks Kepuasan Masyarakat (IKM) meningkat.
- Saran yang direkomendasikan untuk pengembangan penelitian, antara lain:
1. Menggunakan pendekatan *hybrid*, yang merupakan kombinasi *Artificial Neural Network (ANN)* dan *Support Vector Regression (SVR)*.
 2. Pemilihan parameter yang paling tepat dari fungsi kernel dan model *Artificial Neural Network (ANN)* serta diusulkan untuk menggunakan algoritma metaheuristic, *Deep Learning*, dan teknik algoritma dalam akurasi peramalan.
 3. Mencari faktor lain yang dimungkinkan mempengaruhi variabel jumlah izin yang terbit selain dari petugas pelayanan (SDM).

Daftar Rujukan

- [1] P. Meesad and R. I. Rasel., 2013. Predicting Stock Market Price Using Support Regression. *International Conference on Informatics, Electronics and Vision (ICIEV)*.
- [2] Ullu, H. H., 2013. *Prognosis of Damage to Rolling Bearings Using the Support Vector Regression (SVR) Method*. Universitas Diponegoro.
- [3] Hasan, N., Rasel, R. I., and Nath, N. C., 2015. A Support Vector Regression Model for Forecasting Rainfall. *International Conference on Electrical Information and Communication Technologies (EICT)*.
- [4] A. Xu and M. Raginsky., 2017. Information-Theoretic Analysis of Generalization Capability of Learning Algorithm. *31st Conference on Neural Information Processing System (NIPS)*. CA, USA.
- [5] P. Kumar and P. Sharma., 2014. Artificial Neural Network-A Study. *International Journal of Emerging Engineering Research and Technology*, 2 (2), pp. 143-148.
- [6] Nazzal, J. M., El-Emary, I. M., and Najim, S. A., 2008. Multilayer Perceptron Neural Network (MPNN) for Analyzing The Properties of Jordan Oil Shale. *World Applied Sciences Journal*, 5, 546-552.