

MULTI TIME STEPS PREDICTION DENGAN RECURRENT NEURAL NETWORK LONG SHORT TERM MEMORY

Ahmad Ashril Rizal¹, Siti Soraya²

^{1,2}Informatika, STMIK Bumigora Mataram

e-mail: ashril.rizal@stmikbumigora.ac.id, siti.soraya@stmikbumigora.ac.id

Abstrak

Pariwisata telah menjadi sektor andalan dalam pembangunan daerah di Lombok. Kontribusi sektor pariwisata menunjukkan *trend* yang semakin meningkat dari tahun ke tahun. Dampak positif pengeluaran wisatawan terhadap perekonomian terdistribusikan ke berbagai sektor. Akan tetapi, pemerintah daerah umumnya akan melakukan persiapan wisata daerah hanya pada saat *even* lokal saja. Padahal kunjungan wisatawan bukan hanya karena faktor adanya *event* lokal saja. Persiapan pemerintah daerah dan pelaku wisata sangat penting untuk meningkatkan stabilitas kunjungan wisatawan. Penelitian ini mengkaji prediksi kunjungan wisatawan dengan pendekatan *Recurrent Neural Network Long Short Term Memory* (RNN LSTM). LSTM berisi informasi di luar aliran normal dari *recurrent network* dalam *gate cell*. *Cell* membuat keputusan tentang apa yang harus disimpan dan kapan mengizinkan pembacaan, penulisan dan penghapusan, melalui *gate* yang terbuka dan tertutup. *Gate* menyampaikan informasi berdasarkan kekuatan yang masuk ke dalamnya dan akan difilter menjadi bobot dari *gate* itu sendiri. Bobot tersebut sama seperti bobot input dan *hidden* unit yang disesuaikan melalui proses *learning* pada *recurrent network*. Hasil penelitian yang dilakukan dengan membangun model prediksi kunjungan wisatawan dengan RNN LSTM menggunakan *multi time steps* mendapatkan hasil RMSE sebesar 6888.37 pada data *training* dan 14684.33 pada data *testing*.

Keyword — **Sequence to sequence analysis, Recurrent Neural Network, Prediction, Time Series, Tourist Arrivals**

I. PENDAHULUAN

Peranan pariwisata dalam pembangunan ekonomi sangat signifikan. Dengan tidak tersedianya sumber daya alam seperti migas, hasil hutan ataupun industri manufaktur yang berskala besar di provinsi Nusa Tenggara Barat, pada pulau Lombok khususnya, maka pariwisata telah menjadi sektor andalan dalam pembangunan. Kontribusi sektor pariwisata menunjukkan *trend* yang semakin meningkat dari tahun ke tahun. Salah satu yang memberikan kontribusi adalah valuta asing. Penukaran valuta asing dan pengeluaran wisatawan di daerah pariwisata menjadi implikasi yang baik terhadap pendapatan masyarakat.

Pengaruh jumlah wisatawan mancanegara dan lama tinggal secara parsial berpengaruh

positif terhadap penerimaan Produk Domestik Regional Bruto. Selain itu industri pariwisata juga menyumbang peran terhadap pendapatan masyarakat. Jika ditinjau dari kesempatan kerja maka hampir seluruh kesempatan kerja yang ada di Indonesia, Nusa Tenggara pada khususnya, dikontribusikan oleh pariwisata. Hal ini bisa saja terjadi karena kesempatan kerja yang ditimbulkan oleh pengeluaran wisatawan dan akibat investasi di sektor pariwisata. Dampak positif pengeluaran Wisatawan terhadap perekonomian terdistribusikan ke berbagai sektor, bukan saja hotel dan restoran. Distribusi tersebut juga terserap ke sektor pertanian, sektor industri dan kerajinan, sektor pengangkutan dan komunikasi, sektor jasa dan sebagainya. Hal ini sejalan dengan data mengenai distribusi pengeluaran wisatawan.

Data menunjukkan bahwa selama berkunjung, pengeluaran wisatawan yang terserap ke dalam perekonomian rakyat cukup tinggi [1]. Dari riset tersebut digeneralisasi bahwa peran wisatawan memiliki pengaruh positif terhadap perekonomian di Indonesia.

Di sisi lain, Pemerintah Daerah mempersiapkan pariwisata daerah secara optimal hanya pada *event* tertentu. Sementara itu kunjungan wisatawan tidak hanya akan terpusat saat terdapat *event* tertentu. Begitu juga dengan pelaku pariwisata, misalnya dari hotel, penyedia jasa wisata atau akomodasi. Jika kunjungan wisatawan dapat diprediksi maka pelaku pariwisata dapat lebih mempersiapkan diri dengan optimal.

Prediksi adalah salah satu unsur yang sangat penting dalam pengambilan keputusan, sebab efektif atau tidaknya suatu keputusan umumnya tergantung dari beberapa faktor yang tidak dapat kita lihat pada waktu keputusan diambil yang didasarkan pada data yang ada pada waktu sekarang dan waktu lampau [2]. Dalam melakukan prediksi terdapat dua pendekatan yang digunakan. Pendekatan pertama disebut dengan *time-series* yaitu model yang tidak memperlihatkan kecenderungan dari data masa lalu yang tersedia, sedangkan pendekatan yang kedua adalah pendekatan yang memperlihatkan hubungan sebab akibat (*cause-effects method*) atau pendekatan yang menjelaskan terjadinya suatu keadaan (*eksplanatory method*) oleh sebab-sebab tertentu. Permasalahan yang muncul kemudian adalah bagaimana melakukan prediksi. Pada awalnya untuk melakukan prediksi digunakan metode peramalan seperti *Autoregressive Integrated Moving Average Model* (Arima). Metode ini memiliki keterbatasan pada pengabaian kemungkinan hubungan non linear serta stationeritas data dan homokedastitas residual [3].

Kini metode peramalan data dengan *time-series* telah berkembang dengan pendekatan *Neural Network*. Model Arima dan *Neural Network* memiliki perbedaan yaitu Arima baik digunakan untuk meramal data *time series* yang linear sementara *Neural Network* baik digunakan untuk data linear maupun non linear [4].

Penelitian sebelumnya terkait prediksi kunjungan wisatawan menggunakan pendekatan Narma dan Jaringan Saraf Tiruan. Hasil penelitian yang diberikan [5] mengambil kesimpulan bahwa prediksi dengan *time series* memberikan akurasi hasil prediksi yang lebih baik. Penelitian ini mencoba melakukan prediksi dengan melakukan *multi time steps* pada kunjungan wisatawan untuk mengkaji data *time-series* pada kunjungan wisatawan di pulau Lombok dengan menggunakan pendekatan *Recurrent Neural Network* (RNN) *Long Short Term Memory* (LSTM).

Dalam penelitian ini, digunakan beberapa referensi sumber pustaka yang berasal dari penelitian yang sudah dilakukan sebelumnya. Adapun beberapa penelitian mengenai prediksi menggunakan pendekatan *neural network* akan dibahas pada bab ini.

Prediksi data *time-series* dilakukan oleh [4] mengenai model *hybrid* Arima dengan *Neural Network* untuk memprediksi data *time-series*. Model Arima baik digunakan untuk prediksi data *time-series* yang mengandung komponen linear. Hasil penelitian menunjukkan bahwa JST dapat digunakan untuk prediksi data *time-series* dengan komponen non linear. Model *hybrid* dari Arima-NN memiliki tingkat akurasi prediksi yang lebih akurat dibandingkan dengan model Arima tradisional. Sementara itu [6] melakukan pengujian sistem terhadap implementasi metode Arima dan Artificial Neural Network (ANN) serta penggabungan Arima dan ANN. Data yang tidak *stationer* akan *didifferencingkan* terlebih dahulu dengan menggunakan ACF (*Autocorelation Function*) dan PACF (*Partial Autocorelation Function*). Dari hasil penelitian dinyatakan bahwa perhitungan error MSE dan MAPE dengan menggunakan ANN adalah yang paling besar. Nilai MSE dan MAPE saat menggunakan penggabungan Arima dan ANN menunjukkan hasil yang beragam dari tiap inputan. Di tahun berikutnya, [3] melakukan penelitian dengan metode *General Regression Neural Network* (GRNN) untuk memprediksi IHSG. Dari hasil penelitian menghasilkan MSE untuk *insample* dan *outsample* prediksi IHSG berturut-turut adalah 0.0136 dan 0.0135. GRNN menghasilkan nilai MSE prediksi yang lebih kecil

dari Arima. Keunggulan dari GRNN adalah lebih cepat dalam proses perhitungan dan tidak memerlukan adanya asumsi data.

Prediksi lainnya yang menerapkan RNN dilakukan oleh [7] mengenai peramalan harga saham menggunakan RNN dengan Algoritma *Backpropagation Through Time* (BPTT). Peramalan harga saham secara *time-series* dari hasil uji coba yang dilakukan mengenai peramalan harga saham menggunakan RNN-BPTT menghasilkan nilai error yang berbeda-beda. Dengan menerapkan metode *Elman Recurrent Neural Network* (ERNN) dan *Principal Component Analysis* (PCA) [8] melakukan peramalan konsumsi listrik. PCA untuk setiap jenis peramalan sudah dapat diketahui faktor-faktor dominan yang memengaruhi konsumsi listrik dan pemodelan Arima Box Jenkins sudah dapat digunakan untuk menentukan lag-lag input. Hasil pelatihan dengan jaringan dengan menggunakan ERNN untuk setiap jenis peramalan memiliki parameter yang berbeda. Hasil pengujian *forecasting insample* untuk periode peramalan 5 tahun diperoleh nilai rata-rata dari MAPE untuk peramalan konsumsi total 1 sebesar 0.33%, konsumsi total 2 sebesar 0.64%, rumah tangga 1.21%, industri 2.62%, bisnis 3.25%, sosial 0.77% dan publik 0.49%.

Prediksi dengan menerapkan RNN dilakukan [9]. Fokus penelitian ini adalah prediksi secara online, tugas yang dikerjakan jauh lebih sulit daripada inferensi gramatikal dengan jaringan saraf secara offline. Analisis pekerjaan ini penggunaan *discrete-time* RNN dan kemampuan RNN untuk memprediksi simbol berikutnya secara berurutan. [10] melakukan prediksi kecepatan angin dengan menggunakan informasi spasial dari stasiun pengukuran jarak jauh. Metode yang digunakan adalah *Local Recurrent Neural Network*. Untuk meningkatkan ketepatan prediksi, digunakan algoritma pembelajaran online berdasarkan *Recursive Prediction Error* (RPE). Skema RPE dikembangkan dengan semua bobot diperbaharui secara bersamaan. Hasil simulasi ekstensif menunjukkan bahwa model ini menunjukkan hasil prediksi yang lebih baik dibandingkan dengan jenis jaringan lain yang disarankan dalam literatur pada penelitian tersebut.

Penerapan *Neural Network* dengan *Extended*

Kalman Filter telah dilakukan oleh [11] untuk meneliti *Market Clearing Price* (MCP) *Prediction* dengan menerapkan *Multi Layer Perceptron* dengan pelatihan bobot menggunakan EKF. Karena prediksi MCP adalah masalah dengan proses non-stationer maka digunakan EKF sebagai pembelajaran adaptif denebih baik dalam melakukan klasifikasi. Penelitian mengenai prediksi kunjungan Wisatawan sebelugan metode estimasi *convidence* interval. Penelitian yang menerapkan NN-EKF dilakukan juga oleh [12] dalam memprediksi banjir dengan menggunakan ANN dengan EKF. Penelitian ini menekankan pada ANN untuk mendapatkan model yang paling baik dalam akurasi prediksi banjir. Setelah itu di tahun berikutnya [13] melakukan klasifikasi dengan menggunakan *Feedforward Neural Network* dengan EKF. Hasil penelitian memberikan hasil perbandingan dengan klasifikasi menggunakan *gradient descent*. EKF memberikan hasil yang lebih baik dalam melakukan klasifikasi.

Penelitian mengenai prediksi kunjungan Wisatawan sebelumnya dilakukan oleh [2] mengenai prediksi kunjungan Wisatawan Mancanegara di Museum Kota Yogyakarta dengan menerapkan Arima, Algoritma Genetika (AG) dan Neural Network (NN). Data tersebut merupakan data *time-series* yang tidak stasioner sehingga prediksi dengan menggunakan Arima tidak dapat langsung dilakukan, oleh sebab itu harus dilakukan differencing terlebih dahulu. Sementara itu pembentukan kromosom pada model AG menggunakan model Arima ditujukan untuk mendapatkan parameter model yang diambil dari koefisien-koefisien Arima. Pengujian peramalan pada model NN tidak perlu melakukan pengenalan pola data seperti pada Arima. Di tahun berikutnya [14] melakukan penelitian peramalan jumlah kunjungan Wisatawan yang masuk ke dalam suatu daerah di Kusuma Agrowisata Batu Malang. Metode yang digunakan dalam penelitian ini adalah metode Box Jenkins dengan pendekatan model Sarima sebagai pengembangan dari model Arima. Langkah pertama yang dilakukan adalah melihat kestasioneran data. Selanjutnya identifikasi model dari perhitungan ACF dan PACF. Dari perhitungan ACF dan PACF bisa dibentuk model

Arima sementara, kemudian estimasi parameter model dan langkah yang terakhir adalah pemeriksaan diagnostik dengan melihat hasil residual dan normalitas.

[15] mencoba melakukan prediksi *time series* dengan pendekatan *sequence to sequence* pada kunjungan wisatawan. Arsitektur dibangun dengan *Recurrent Neural Network* dan *training RNN* dilakukan dengan *Extended Kalman Filter*. Di tahun berikutnya [5] melakukan prediksi dengan pendekatan Narma pada kunjungan wisatawan. Hasil penelitiannya menunjukkan bahwa RNN baik untuk melakukan *sequence to sequence analysis*, sementara Narma baik untuk prediksi dengan *multivariate* dimana terdapat beberapa faktor yang mempengaruhi kunjungan wisatawan.

Penelitian ini mencoba untuk melakukan prediksi kunjungan wisatawan dengan pendekatan *multi time steps* menggunakan RNN LSTM (*Long Short Term Memory*).

II. METODOLOGI

Sistem yang dibuat untuk memprediksi jumlah kunjungan Wisatawan dengan menggunakan *Recurrent Neural Network*. Proses prediksi kunjungan Wisatawan diawali dengan tahap penentuan data input yang digunakan untuk prediksi. Variabel yang digunakan adalah jumlah kunjungan Wisatawan perbulan dari Januari 2009 hingga Desember 2014. Data input tersebut diberikan proses pembelajaran pada jaringan *Recurrent Neural Network* dengan *Long Short Term Memory* untuk melakukan prediksi. Hasil *training* akan diuji dengan data *testing* untuk menentukan validasi hasil prediksi.

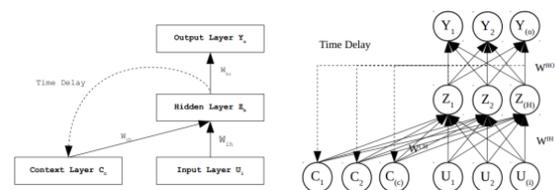
2.1 Rancangan Sistem Prediksi kunjungan

Dalam merancang model prediksi dengan *Recurrent Neural Network* (RNN) *Long Short Term Memory* (LSTM) dilakukan beberapa tahapan yaitu perancangan arsitektur, menentukan inisialisasi bobot awal, inisialisasi parameter, vektor input, vektor output dan penetapan target jaringan. Arsitektur RNN dibangun dalam bentuk LSTM sebagai bagian dari pembelajaran terawasi. Bobot dalam RNN dilatih dengan menggunakan LSTM.

2.2.1 Arsitektur *Recurrent Neural Network*

Recurrent Neural Network merupakan jaringan saraf berulang. Dikatakan jaringan saraf berulang karena nilai *neuron* pada *hidden layer* sebelumnya akan digunakan kembali sebagai data input. Penggunaan *neuron* pada *hidden layer* akan disimpan ke dalam sebuah *layer* yang dinamakan *context layer*. Nilai *neuron* pada *context layer* akan terus *update* hingga kondisi RNN terpenuhi.

Gambar 1 menunjukkan arsitektur RNN yang digunakan. Jumlah *neuron* dalam input *layer* adalah lima buah *neuron* dan sebuah *neuron* bias. Penambahan bias pada *neuron* di input *layer* dan *hidden layer* diberikan untuk membantu proses pembelajaran. Pada dasarnya penambahan bias akan menambah beban perhitungan namun bisa membantu pergerakan *neuron* pada ambang fungsi aktivasi yang digunakan.



Gambar 1. Arsitektur RNN

Dalam tulisan ini jumlah *neuron* pada *hidden layer* dan *context layer* yang digunakan adalah sama yaitu sejumlah *k neuron*. Jumlah *neuron* yang sama pada *hidden layer* dengan *context layer* mengikuti arsitektur yang menerapkan Elman *Recurrent Neural Network*.

2.3 Proses *Training* dengan RNN-EKF

2.3.1 Data *Training*

Data *training* adalah data yang digunakan selama proses *training* dalam sistem RNN-EKF. Dari total dataset yang ada, persentase data *training* adalah 70% dari total data tersebut. Data *training* yang masuk ke dalam jaringan sebanyak 300 titik data. Tiap *neuron* dalam input *layer* mewakili vektor input yang melibatkan 60 data *training*.

2.3.2 Preprocessing

Preprocessing dalam penelitian meliputi normalisasi data *training*. Sebelum diproses data-data input tersebut akan dinormalisasi. Normalisasi data dilakukan agar keluaran jaringan sesuai dengan fungsi aktivasi yang digunakan. Data-data tersebut dinormalisasi ke dalam interval $[-1,1]$. Interval tersebut merupakan batas nilai untuk fungsi aktivasi *hyprbolic tangent*.

Dalam arsitektur RNN yang digunakan terdapat tiga jenis bobot yang digunakan. Nilai dari bobot awal dari RNN yang dibangun adalah nilai *random* dari pada interval $[0,10]$.

Adapun bobot-bobot dalam RNN yang digunakan adalah bobot dari input layer ke *hidden layer* (w_{ih}), bobot dari *hidden layer* ke *output layer* (w_{ho}) dan bobot dari *context layer* ke *hidden layer* yang diinisialisasi dengan w_{ch} . Jumlah node c sama dengan node h karena dalam arsitektur yang digunakan jumlah *neuron* pada *hidden layer* sama dengan jumlah *neuron* pada *context layer*.

2.4 Long Short Term Memory Proses

Setelah data diperoleh, kemudian berdasarkan teori-teori yang ada dan dengan menggunakan metode analisis time series, maka dilakukan proses prediksi kunjungan wisatawan. Model prediksi yang digunakan adalah RNN LSTM. Langkah dalam proses prediksi dengan menggunakan metode RNN LSTM adalah sebagai berikut:

1. Input Dataset

Dataset berisi data *training* dan data *testing*. Data *training* adalah data yang digunakan selama proses *training* dalam sistem RNN LSTM. Dari total dataset yang ada, persentase data *training* adalah 70% dari total dataset. Tiap *neuron* dalam input layer mewakili vektor input yang melibatkan data *training*. Data *training* disimpan dalam sebuah file csv (*comma-separated values*).

2. Normalisasi

Preprocessing dalam penelitian meliputi normalisasi data *training*. Sebelum diproses, data-data input tersebut akan dinormalisasi. Normalisasi terhadap data dilakukan agar keluaran jaringan sesuai dengan fungsi aktivasi yang digunakan. Data-data tersebut

dinormalisasi ke dalam interval $[-1,1]$. Interval tersebut merupakan batas nilai untuk fungsi aktivasi *hyprbolic tangent*. Normalisasi dapat dilakukan dengan persamaan (1).

$$n_i = \frac{2(x_i - x_{min})}{x_{max} - x_{min}} - 1 \quad (1)$$

Keterangan:

n_i = Data hasil normalisasi

x_i = Data ke-i

x_{min} = Data dengan nilai minimum

x_{max} = Data dengan nilai maksimum

3. Desain Multi Time Steps

Desain *multi time steps* dilakukan untuk melihat *sequence* dari data yang digunakan. Desain *multi time steps* akan berpengaruh terhadap arsitektur yang akan digunakan dalam sistem. Model desain akan menentukan model input dan output dari RNN LSTM yang digunakan.

4. Neural Network Process

Dalam arsitektur RNN LSTM terdapat tiga jenis bobot yang digunakan. Adapun bobot-bobot dalam RNN yaitu bobot dari *input layer* ke *hidden layer*, bobot dari *hidden layer* ke *output layer* dan bobot dari *context layer* ke *hidden layer*. Selain bobot, proses RNN LSTM juga termasuk inisialisasi parameter pembelajaran. Parameter pembelajaran dalam RNN LSTM pada penelitian ini adalah nilai *learning rate*, *epoch* dan fungsi aktivasi. Proses *training* akan dihentikan ketika nilai *error* memenuhi target atau iterasi maksimal yang telah ditetapkan terpenuhi.

5. Denormalisasi

Sebelum menghitung akurasi hasil prediksi, terlebih dahulu dilakukan proses denormalisasi hasil output dari RNN LSTM. Denormalisasi dilakukan untuk mendapatkan nilai real dari hasil prediksi yang diberikan. Sementara itu akurasi dihitung dengan melihat persen akurasi dari hasil prediksi. Proses normalisasi dilakukan dengan persamaan (2).

$$dn_i = ((n_i + 1)(x_{ma} - x_{min})) + \left(\frac{2(x_{min})}{2}\right) \quad (2)$$

Keterangan:

dn_i = Data hasil normalisasi

n_i = Data ke- i

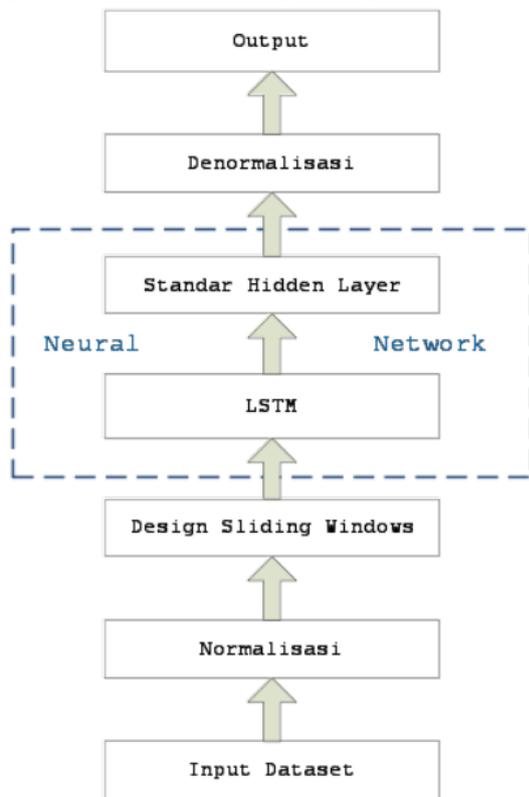
x_{min} = Data dengan nilai minimum

x_{max} = Data dengan nilai maksimum

6. Output

Output dari sistem adalah hasil prediksi dan akurasi hasil prediksi. Hasil akurasi prediksi dihitung dengan melihat persentase akurasi.

Adapun langkah prediksi data *time series* dengan RNN LSTM ditunjukkan pada Gambar 3.

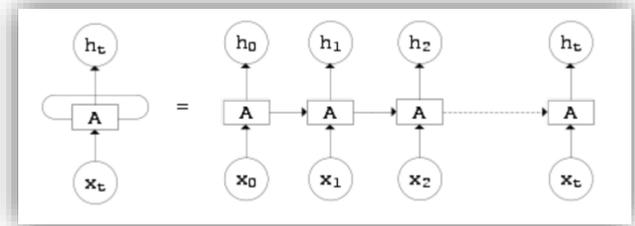


Gambar 3. Flowchart Sistem Prediksi dengan LSTM

2.4.2 Desain Prediksi Time Series dengan Long Short Term Memory

Sistem yang dibangun dalam penelitian ini adalah sistem prediksi kunjungan wisatawan di Pulau Lombok menggunakan *Recurrent Neural Network Longs Short Term Memory*. Arsitektur RNN LSTM yang digunakan ditunjukkan pada Gambar

4. Pada jaringan, algoritma yang digunakan untuk *training* adalah *Backpropagation Through Time*.



Gambar 4. Arsitektur LSTM

Sebelum memulai pelatihan, terlebih dahulu ditentukan arsitektur dan parameter jaringan. Selain itu ditentukan juga jumlah neuron pada *input layer*, *hidden layer* dan *output layer*. Output hasil prediksi dari sistem berupa single output yang diharapkan sesuai dengan kondisi sesungguhnya. Akurasi prediksi akan ditentukan oleh besarnya error pada saat *testing*.

III. HASIL DAN PEMBAHASAN

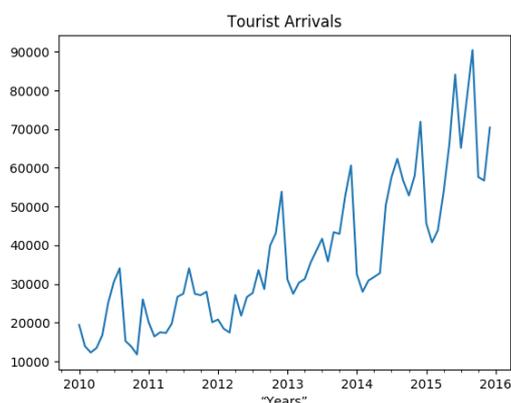
Masalah prediksi *time series* adalah salah satu jenis masalah yang rumit untuk pemodelan prediksi. Berbeda dengan pemodelan prediksi dengan regresi, prediksi *time series* berisi sekumpulan variabel input. Jenis *neural network* yang kuat dirancang untuk menangani *sequence dependences* disebut *recurrent neural network* (jaringan saraf berulang). *Long Short Term Memory* (LSTM) adalah salah satu jenis *recurrent neural network* yang digunakan dalam pembelajaran mendalam karena arsitekturnya yang kompleks dan dapat dilatih.

Penelitian ini mengembangkan jaringan LSTM dengan menggunakan bahasa pemrograman python. Penelitian ini mengkaji masalah prediksi *time series* pada kunjungan wisatawan di pulau Lombok. Selain itu, penelitian ini juga mengembangkan jaringan LSTM untuk regresi, *sliding window* berdasarkan waktu dari masalah prediksi *time series*. LSTM membuat prediksi dengan mempertahankan *state* (memori) di *sequence* yang sangat panjang.

Masalah yang akan dibahas dalam penelitian ini adalah masalah prediksi kunjungan wisatawan mancanegara di pulau Lombok. Masalah dalam

kasus ini adalah memprediksi kunjungan wisatawan pada bulan berikutnya dengan data yang diberikan tiap bulan dalam tiap tahun. Data dimulai dari Januari 2010 hingga Desember 2016. Dengan demikian dataset adalah 7 tahun atau 84 pengamatan. Kumpulan data ini diperoleh dari Dinas Pariwisata NTB dan Badan Pusat Statistik NTB. File dataset disimpan dalam format csv atau *comma separated value*.

Dataset akan diakuisisi dengan bantuan library *pandas* dari *python*. Data tanggal akan diabaikan, mengingat bahwa setiap pengamatan dipisahkan oleh interval yang sama dalam satu bulan. Oleh karena itu dataset akan dimuat dengan mengecualikan data pada kolom pertama. Gambar 5 adalah grafik himpunan dataset yang digunakan dalam pengamatan.



Gambar 5. Grafik Dataset Kunjungan Wisatawan

Dalam penelitian ini, sistem akan tetap membuat data sederhana dan apa adanya. Meskipun pada umumnya, prediksi dilakukan dengan cara itu perbaikan data menggunakan teknik *rescale* dan membuat data stasioner.

5 series data training awal adalah sebagai berikut:

Month	Tourist Arrivals
2010-01-01	19435
2010-02-01	13931
2010-03-01	12292
2010-04-01	13483
2010-05-01	16781

Name: #TouristArrivals, dtype: int64

3.1 Multi Time Steps Predictions

Dari dataset akan dibuat prediksi dengan merancang dataset menjadi *multi time steps*. Untuk bulan tertentu dalam 12 bulan terakhir kumpulan data, akan dibuat time step dalam kurun waktu 3 bulan. Atau dengan kata lain, diberikan pengamatan historis ($t-1, t-2, \dots, t-n$) ramalan $t, t+1$ dan $t+2$. Secara spesifik, untuk meramalkan kondisi di bulan April 2018, akan diramalkan dengan data Januari, Februari dan Maret. Untuk bulan Mei, diramalkan dengan data pada bulan Februari, Maret dan April. Sebagai contoh 10 prakiraan dengan data 3 bulan seelumnya, sebagai berikut:

Dec, Jan, Feb, Mar
Jan, Feb, Mar, Apr
Feb, Mar, Apr, May
Mar, Apr, May, Jun
Apr, May, Jun, Jul
May, Jun, Jul, Aug
Jun, Jul, Aug, Sep
Jul, Aug, Sep, Oct
Aug, Sep, Oct, Nov
Sep, Oct, Nov, Dec

3.2 Evaluasi Model

Skenario *rolling-forecast* yang akan digunakan juga disebut dengan validasi model *walk-forward*. Setiap langkah waktu dari dataset testing akan berjalan satu per satu. Model akan digunakan untuk membuat perkiraan untuk langkah waktu, maka nilai yang diharapkan aktual untuk bulan depan dari rangkaian tes akan diambil dan tersedia untuk model perkiraan pada langkah waktu berikutnya.

Semua perkiraan pada kumpulan data uji akan dikumpulkan dan skor kesalahan dihitung untuk merangkum ketrampilan model untuk setiap langkah waktu perkiraan. Root mean squared error (RMSE) akan digunakan karena menghukum kesalahan besar dan menghasilkan skor yang sama dengan data perkiraan, yaitu kunjungan Wisatawan.

Baseline yang baik untuk peramalan *time series* adalah model distribusi normal. Ini adalah model peramalan dimana pengamatan terakhir terus berlanjut. Karena kesederhanaannya, sering

disebut ramalan naif.

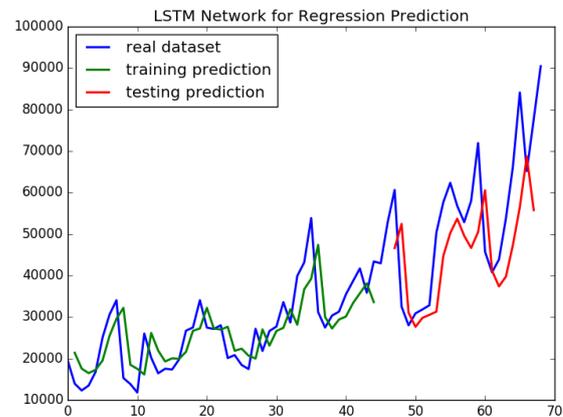
3.3 Long Short-Term Memory Regression Prediction

LSTM dapat dibawa ke dalam masalah regresi linear. Prediksi dengan regresi akan menganalisa jumlah jumlah kunjungan pada bulan t untuk memprediksi kunjungan pada bulan ke $t+1$. Fungsi sederhana akan diberikan untuk mengonversi kolom tunggal pada dataset menjadi kumpulan data dua kolom. Di mana kolom pertama berisi jumlah kunjungan pada bulan ini (t) dan kolom kedua berisi jumlah kunjungan penumpang bulan depan ($t+1$).

Sebelum memulai prediksi, dilakukan ekstraksi array dari *dataframe* dan mengkonversi nilai integer ke nilai *floating point* yang lebih cocok dengan model jaringan syaraf yang digunakan. LSTM sangat sensitif terhadap skala data input, khususnya ketika fungsi aktivasi sigmoid (default) atau tanh digunakan. Pada penelitian ini digunakan fungsi aktivasi tanh. Fungsi aktivasi tersebut melakukan skala ulang data pada ke kisaran -1 hingga 1, juga disebut normalisasi. Normalisasi dataset adalah bagian dari preprocessing dengan menerapkan *MinMaxScaler* dari *scikit-learn*.

Setelah memodelkan data akan dilakukan validasi menggunakan *cross over validation*. Data *time series* yang dijadikan dataset sangat memperhatikan urutan. Metode sederhana yang dapat digunakan yaitu dengan membagi dataset menjadi data uji dan data test. Dalam penelitian ini, dataset yang dijadikan data *training* sebesar 67% pengamatan dan dataset yang dijadikan data *testing* adalah 33%. Jaringan memiliki 1 input layer, 1 hidden layer dengan 4 blok LSTM atau neuron dan output layer yang akan memberikan prediksi tunggal. Fungsi aktivasi sigmoid default digunakan untuk blok LSTM. Jaringan dilatih untuk 100 iterasi dengan *batch size* 1. Error yang digunakan sebagai stop kriteria adalah *Root Mean Square Error* (RMSE). Pada saat memberikan hasil prediksi, hasil prediksi akan digeser agar sejajar pada sumbu x dengan dataset asli. Setelah disiapkan, data diplot, dataset asli akan direpresentasikan dengan warna biru, prediksi untuk dataset pelatihan dalam warna hijau dan prediksi pada dataset uji dalam warna merah.

Hasil prediksi dengan LSTM regresi ditunjukkan pada Gambar 6.



Gambar 6. Hasil Prediksi LSTM Regresi

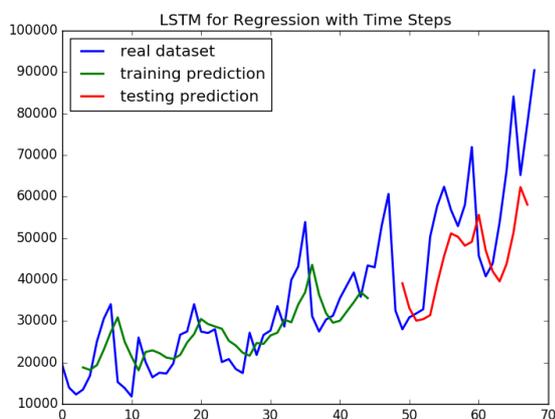
Hasil *training* dengan menggunakan LSTM regresi jika ditinjau dari nilai RMSE adalah 6529,42. Sementara itu hasil RMSE pada data *testing* adalah 13766,85.

3.4 LSTM for Regression with Multi Time Steps

Dalam metode ini, akan diterapkan proyeksi bergulir atau disebut dengan model validasi *walk-forward*. Setiap time step dari test dataset akan berjalan di saat yang bersamaan. Sebuah model akan digunakan untuk membuat perkiraan satu waktu ke depan. Nilai sebenarnya yang diharapkan dari test dataset akan diambil dan digunakan sebagai model untuk perkiraan langkah waktu berikutnya. Hal ini sama dengan prediksi secara konvensional di mana pengamatan kunjungan wisatawan akan tersedia setiap bulan dan akan digunakan untuk perkiraan bulan berikutnya. Prediksi ini akan diuji pada data *training* dan data *testing*.

Gambar 7 menunjukkan hasil prediksi dengan LSTM *sliding window*. Sebelum memasukkan model LSTM ke dataset, data harus diproses terlebih dahulu. Transformasi data dilakukan pada dataset sebelum dimasukkan ke dalam model untuk membuat perkiraan. Transformasikan data *time series* yang pertama ditunjukkan agar data stasioner. Secara khusus, lag=1 akan melakukan *differencing* untuk menghapus kecenderungan

data menaik (*increasing trend*). Transformasi *time series* berikutnya untuk mengubah data menjadi masalah pembelajaran yang diawasi



Gambar 7. Grafik Prediksi dengan *Time Step*

Organisasi data menjadi input dan output di mana pengamatan pada langkah waktu sebelumnya digunakan sebagai input untuk meramalkan observasi pada waktu saat ini. Transformasi pengamatan untuk mendapatkan skala sesuai kebutuhan aktivasi. pada penelitian ini dilakukan rescale data pada nilai antara -1 dan 1 untuk memenuhi fungsi aktivasi tangen hiperbolic tangent dari model LSTM. Transformasi ini akan diberikan proses inversi setelah proses prediksi untuk mengembalikan nilai ke skala aslinya sebelum menghitung dan skor kesalahan. Hasil prediksi dengan time step ditunjukkan pada Gambar 8 dengan RMSE sebesar 6888.37 pada data training dan 14684.33 pada data *testing*.

IV. SIMPULAN DAN SARAN

Pada penelitian ini telah dilakukan prediksi data *time series* dengan kasus kunjungan wisatawan di pulau Lombok. Berdasarkan hasil penelitian yang dilakukan dengan membangun sistem prediksi kunjungan wisatawan dengan *Recurrent Neural Network Long Short Term Memory* (RNN LSTM), diperoleh kesimpulan bahwa training RNN LSTM dengan menggunakan tiga model memberikan hasil yang beragam. Dari ketiga model LST yang dilakukan yaitu LSTM regression, LSTM dengan sliding window dan LSTM dengan *time steps* tidak ada model yang

memberikan hasil optimal dari sisi *training* dan *testing* sekaligus. Hasil terbaik pada proses training untuk prediksi kunjungan wisatawan diperoleh dengan menggunakan model regresi dengan RMSE 6529,42.

Penelitian dapat dilakukan dengan mengeksplorasi algoritma *training* yang lain untuk prediksi dengan menggunakan jaringan *recurrent neural network*. Dengan metode *training* yang berbeda dimungkinkan kekurangan yang terdapat prediksi dengan RNN memberikan hasil dengan tingkat akurasi yang lebih baik. Selain itu, penelitian terkait optimasi penentuan bobot awal sehingga proses *training* dapat berjalan lebih cepat.

V. UCAPAN TERIMA KASIH

Atas selesai hingga terbitnya artikel ini, kami menyampaikan ucapan terima kasih dan penghargaan yang setinggi-tingginya kepada semua anggota tim peneliti serta Direktorat Riset dan Pengabdian Masyarakat (DRPM) yang telah memberikan hibah Penelitian Dosen Pemula.

REFERENSI

- [1] I. N. Wijaya, "Pengaruh Jumlah Wisatawan Mancanegara, Lama Tinggal, Dan Kurs Dollar Amerika terhadap Penerimaan Produk Domestik Regional Bruto Industri Pariwisata di Badung," 2013.
- [2] F. A. Setyaningsih, "PERBANDINGAN ALGORITMA GENETIKA DAN METODE STATISTIK ARIMA UNTUK PREDIKSI DATA TIME SERIES (Studi Kasus: Kunjungan Museum di Kota Yogyakarta)," in *Prosiding Semirata 2015 Bidang Teknologi Informasi dan Multi Disiplin*, 2015, pp. 69–82.
- [3] L. P. W. Adnyani and Subanar, "General Regression Neural Network (GRNN) Pada Peramalan Kurs Dolar Dan Indeks Harga Saham Gabungan (IHSG)," *Factor Exacta*, vol. 8, no. 2, pp. 137–144, 2015.
- [4] E. Munarsih, "PENERAPAN MODEL ARIMA-NEURAL NETWORK HYBRID UNTUK PERAMALAN TIME SERIES,"

- Universitas Gadjah Mada, 2011.
- [5] S. H. Ahmad Ashril Rizal, "Prediksi kunjungan wisatawan dengan recurrent neural network extended kalman filter," *Ilmu Komputer, Udayana*, vol. I, no. June, p. 1, 2017.
- [6] A. Rumagit, S.E., "Prediksi Pemakaian Listrik Menggunakan Jaringan Syaraf Tiruan dan ARIMA di Wilayah Sullutenggo," *IndoCEISS*, vol. 7, no. 2, pp. 139–148, 2013.
- [7] L. A. D. Susanti, A. Fariza, and Setiawardhana, "Peramalan Harga Saham Menggunakan Recurrent Neural Network Dengan Algoritma Backpropagation Through Time," *PENS-ITS*, no. January, pp. 1–8, 2011.
- [8] T. Rahmawati, "Perancangan Augmented Reality Volcano untuk Alat Peraga Museum," *Teknomatika*, vol. 5, pp. 1–11, 2013.
- [9] J. A. Pérez-Ortiz, J. Calera-Rubio, and M. L. Forcada, "Online text prediction with recurrent neural networks," *Neural Process. Lett.*, vol. 14, no. 2, pp. 127–140, 2001.
- [10] T. G. Barbounis and J. B. Theocharis, "Locally recurrent neural networks for wind speed prediction using spatial correlation," *Inf. Sci. (Ny)*, vol. 177, no. 24, pp. 5775–5797, 2007.
- [11] L. Zhang and P. B. Luh, "Neural network-based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 59–66, 2005.
- [12] R. Adnan, F. A. Ruslan, A. M. Samad, and Z. M. Zain, "New Artificial Neural Network and Extended Kalman Filter hybrid model of flood prediction system," *Signal Process. its Appl. (CSPA), 2013 IEEE 9th Int. Colloq.*, pp. 252–257, 2013.
- [13] A. N. Chernodub, "Training Neural Networks for classification using the Extended Kalman Filter: A comparative study," *Opt. Mem. Neural Networks*, vol. 23, no. 2, pp. 96–103, 2014.
- [14] N. Lestari *et al.*, "Peramalan Kunjungan Wisata dengan Pendekatan Model SARIMA (Studi kasus : Kusuma Agrowisata)," vol. 1, no. 1, 2012.
- [15] A. A. Rizal and S. Hartati, "Recurrent neural network with Extended Kalman Filter for prediction of the number of tourist arrival in Lombok," in *2016 International Conference on Informatics and Computing (ICIC)*, 2016, pp. 180–185.