

TESIS

**PREDIKSI NILAI REALISASI INVESTASI MENGGUNAKAN
METODE SUPPORT VECTOR REGRESSION (SVR) DAN
LONG SHORT TERM MEMORY (LSTM)
(Studi Kasus: DPMPSTP Kota Magelang)**



Disusun oleh:

Nama : Christina Andriyani
NIM : 20.52.1313
Konsentrasi : Informatics Technopreneurship

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

TESIS

**PREDIKSI NILAI REALISASI INVESTASI MENGGUNAKAN
METODE SUPPORT VECTOR REGRESSION (SVR) DAN
LONG SHORT TERM MEMORY (LSTM)
(STUDI KASUS : DPMPSTP KOTA MAGELANG)**

**PREDICTION OF INVESTMENT REALIZATION VALUE USING
SUPPORT VECTOR REGRESSION (SVR) AND
LONG SHORT TERM MEMORY (LSTM) METHODS
(CASE STUDY : DPMPSTP CITY OF MAGELANG)**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Christlna Andriyani
NIM : 20.52.1313
Konsentrasi : Informatics Technopreneurship

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

HALAMAN PENGESAHAN

**PREDIKSI NILAI REALISASI INVESTASI MENGGUNAKAN
METODE SUPPORT VECTOR REGRESSION (SVR) DAN
LONG SHORT TERM MEMORY (LSTM)
(STUDI KASUS : DPMPSTP KOTA MAGELANG)**

**PREDICTION OF INVESTMENT REALIZATION VALUE USING
SUPPORT VECTOR REGRESSION (SVR) AND
LONG SHORT TERM MEMORY (LSTM) METHODS
(CASE STUDY : DPMPSTP CITY OF MAGELANG)**

Dipersiapkan dan Disusun oleh

Christina Andriyani

20.52.1313

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Senin, 3 Januari 2022

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 3 Januari 2022

Rektor

Prof. Dr. M. Suyanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

**PREDIKSI NILAI REALISASI INVESTASI MENGGUNAKAN
METODE SUPPORT VECTOR REGRESSION (SVR) DAN
LONG SHORT TERM MEMORY (LSTM)
(STUDI KASUS : DPMPSTP KOTA MAGELANG)**

**PREDICTION OF INVESTMENT REALIZATION VALUE USING
SUPPORT VECTOR REGRESSION (SVR) AND
LONG SHORT TERM MEMORY (LSTM) METHODS
(CASE STUDY : DPMPSTP CITY OF MAGELANG)**

Dipersiapkan dan Disusun oleh

Christina Andriyani

20.52.1313

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Senin, 3 Januari 2022

Pembimbing Utama

Prof. Dr. Ema Utami, S. Si, M.Kom
NIK. 190302037

Pembimbing Pendamping

Hanif Al Fatta, M.Kom
NIK. 190302096

Anggota Tim Penguji

Prof Dr. Kusriani, M.Kom
NIK. 190302106

Dr. Andi Sunvoto, M. Kom
NIK. 190302052

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 3 Januari 2022
Direktur Program Pascasarjana

Prof Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Christina Andriyani
NIM : 20.52.1313
Konsentrasi : Informatics Technopreneurship

Menyatakan bahwa Tesis dengan judul berikut:
**Prediksi Nilai Realisasi Investasi Menggunakan Metode Support Vector
Regression (SVR) dan Long Short Term Memory (LSTM)**
(Studi Kasus : DPMPSTP Kota Magelang)

Dosen Pembimbing Utama : Prof. Dr. Erna Utami, S. Si, M.Kom
Dosen Pembimbing Pendamping : Hanif Al Fatta, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta,
Yang Menyatakan,



Christina Andriyani

HALAMAN PERSEMBAHAN

Kupersembahkan Tesis ini untuk keluargaku tercinta :

- Kedua orang tuaku
- Suamiku
- Kedua anakku tercinta Kefas Rama Antoro dan Hana Khariessa Antoro



HALAMAN MOTTO

Apapun juga yang kamu perbuat, perbuatlah dengan segenap hatimu seperti untuk Tuhan bukan untuk manusia. (Kolose 3:23)

Segala hal yang kita lakukan dengan niat dan untuk tujuan yang baik harus kita lakukan dengan sungguh-sungguh dan kita persembahkan sebagai ucapan syukur atas semua anugerah yang diberikan Tuhan kepada kita.



KATA PENGANTAR

Puji dan syukur kepada Tuhan Allah Yang Maha Esa yang telah memberikan karunia dan rahmatNya sehingga penulis dapat menyelesaikan penelitian tesis dengan baik. Pada kesempatan ini penulis hendak mengucapkan terima kasih kepada pihak-pihak yang telah mendukung selama mengerjakan penelitian tesis, antara lain :

1. Prof. Dr. Ema Utami, S.Si., M.Kom dan Hanif Al Fatta, M.Kom yang penuh kesabaran membimbing, memberi saran terbaik dan motivasi untuk penyelesaian tesis ini.
2. Direktur Program Studi Pascasarjana Universitas AMIKOM, Ketua Program Studi Pascasarjana, para dosen program studi S2 Informatics Techopreneurship dan staf admisi MTI AMIKOM.
3. Kedua orangtua, yang selalu memberikan dukungan, semangat dan doa yang tak terbatas untuk kelancaran dan selesainya semua tugas dan tesis.
4. Suami dan anak-anak tercinta yang telah memberikan semangat, dukungan, doa, cinta dan kesabaran selama masa studi.
5. Kepala Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu Kota Magelang Bapak Muchamad Abdul Azis, SH dan rekan-rekan kerja yang mendukung penulis untuk melakukan penelitian dan selalu memberikan semangat dan bantuan dalam bentuk apapun.
6. Teman – teman MTI Universitas AMIKOM Yogyakarta atas persahabatan, semangat dan dukungan dalam bentuk apapun.

Akhirnya kritik dan saran sangat penulis harapkan untuk perbaikan dimasa mendatang. Semoga hasil penelitian ini dapat memberikan manfaat dan pengetahuan bagi banyak pihak.

Yogyakarta, 3 Januari 2022

Penulis



DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS.....	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	10
1.3. Batasan Masalah.....	11
1.4. Tujuan Penelitian.....	11
1.5. Manfaat Penelitian.....	12
BAB II TINJAUAN PUSTAKA.....	13
2.1. Tinjauan Pustaka.....	13
2.2. Landasan Teori.....	20

2.3. Keaslian Penelitian.....	34
BAB III METODE PENELITIAN.....	37
3.1. Jenis, Sifat dan Pendekatan Penelitian.....	37
3.2. Metode Pengumpulan Data.....	37
3.3. Metode Analisis Data.....	38
3.4. Alur Penelitian.....	38
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	41
4.1. Gambaran Umum Obyek Penelitian.....	41
4.2. Pengumpulan Data.....	42
4.3. Implementasi Pembentukan Model.....	48
4.4. Pengujian Model Prediksi.....	54
BAB V PENUTUP.....	89
5.1. Kesimpulan.....	89
5.2. Saran.....	90
DAFTAR PUSTAKA.....	92

DAFTAR TABEL

Tabel 1.1. Perkembangan Penanaman Modal Tahun 2014 – 2019.....	2
Tabel 1.2. Perkembangan Penanaman Modal Tahun 2014 – 2019.....	3
Tabel 2.1. Matriks literatur review dan posisi penelitian prediksi realisasi investasi menggunakan metode Support Vector Regression (SVR) dan Long Short Term Memory (LSTM).....	34
Tabel 4.1. Daftar Atribut Data Realisasi Investasi.....	43
Tabel 4.2. Rentang nilai parameter pada tahap loose grid.....	50
Tabel 4.3. Rentang nilai parameter tahap finer grid skenario bulanan.....	51
Tabel 4.4. Rentang nilai parameter tahap finer grid skenario tahunan.....	51
Tabel 4.5. Rentang nilai parameter tahap finer grid skenario tiga tahunan.....	51
Tabel 4.6. Nilai Parameter terbaik pada seluruh skenario.....	52
Tabel 4.7. Parameter model prediksi algoritma SVR skenario bulanan.....	52
Tabel 4.8. Parameter model prediksi algoritma SVR skenario tahunan.....	52
Tabel 4.9. Parameter model prediksi algoritma SVR skenario tiga tahunan.....	53
Tabel 4.10. Parameter model prediksi algoritma LSTM.....	53
Tabel 4.11. Dataset Skenario Bulanan.....	55
Tabel 4.12. Normalisasi Data Bulanan.....	55
Tabel 4.13. Jarak Antar Data Training Bulanan RBF.....	56
Tabel 4.14. Jarak Antar Data Training Bulanan Polynomial.....	56
Tabel 4.15. Matrik Hessian Data Training Bulanan RBF.....	57
Tabel 4.16. Matrik Hessian Data Training Bulanan Polynomial.....	57

Tabel 4.17 Nilai Error Bulanan.....	58
Tabel 4.18 Perubahan Nilai Langgrange Multipler Bulanan.....	58
Tabel 4.19 Update Perubahan Nilai Langgrange Multipler Bulanan.....	59
Tabel 4.20 Sequential Learning 1 Iterasi Bulanan.....	59
Tabel 4.21 Sequential Learning 100 Iterasi Bulanan.....	59
Tabel 4.22 Data Testing Bulanan.....	60
Tabel 4.23 Jarak Antar Data Testing Bulanan RBF.....	60
Tabel 4.24 Jarak Antar Data Testing Bulanan Polynomial.....	61
Tabel 4.25 Matrik Hessian Testing Bulanan RBF.....	61
Tabel 4.26 Matrik Hessian Testing Bulanan Polynomial.....	61
Tabel 4.27 Hasil Prediksi Bulanan.....	62
Tabel 4.28 Hasil pengujian RMSE algoritma SVR & LSTM- Bulanan.....	65
Tabel 4.29 Dataset Skenario Tahunan.....	68
Tabel 4.30 Normalisasi Data Tahunan.....	68
Tabel 4.31 Jarak Antar Data Training Tahunan RBF.....	69
Tabel 4.32 Jarak Antar Data Training Tahunan Polynomial.....	69
Tabel 4.33 Matrik Hessian Data Training Tahunan RBF.....	70
Tabel 4.34 Matrik Hessian Data Training Tahunan Polynomial.....	70
Tabel 4.35 Nilai Error Tahunan.....	71
Tabel 4.36 Perubahan Nilai Langgrange Multipler Tahunan.....	71
Tabel 4.37 Update Perubahan Nilai Langgrange Multipler Tahunan.....	72
Tabel 4.38 Sequential Learning 1 Iterasi Tahunan.....	72
Tabel 4.39 Sequential Learning 100 Iterasi Tahunan.....	72

Tabel 4.40 Data Testing Tahunan.....	73
Tabel 4.41 Jarak Antar Data Testing Tahunan RBF.....	73
Tabel 4.42 Jarak Antar Data Testing Tahunan Polynomial.....	74
Tabel 4.43 Matrik Hessian Testing Tahunan RBF.....	74
Tabel 4.44 Matrik Hessian Testing Tahunan Polynomial.....	74
Tabel 4.45 Hasil Prediksi Tahunan.....	75
Tabel 4.46 Hasil pengujian RMSE algoritma SVR & LSTM - Tahunan.....	79
Tabel 4.47 Dataset Skenario Tiga Tahunan.....	81
Tabel 4.48 Nilai Koefisien Lagrange.....	82
Tabel 4.49 Hasil Prediksi Tiga Tahunan.....	82
Tabel 4.50 Hasil pengujian RMSE algoritma SVR & LSTM - Tiga tahun.....	85

DAFTAR GAMBAR

Gambar 2.1	Identifikasi Pencilan Menggunakan IQR.....	23
Gambar 2.2	Arsitektur LSTM.....	29
Gambar 3.1	Alur Penelitian.....	40
Gambar 4.1	Nilai korelasi antar Atribut.....	44
Gambar 4.2	Dataset Awal Nilai Realisasi Investasi.....	45
Gambar 4.3	Statisik Dataset Realisasi Investasi.....	46
Gambar 4.4	Dataset Realisasi Investasi.....	47
Gambar 4.5	Arsitektur LSTM Prediksi Bulanan.....	63
Gambar 4.6	Data training dan data testing - SVR bulanan.....	66
Gambar 4.7	Data training dan data testing - LSTM bulanan.....	67
Gambar 4.8	Arsitektur LSTM Prediksi Tahunan.....	76
Gambar 4.9	Data training dan data testing - SVR tahunan.....	80
Gambar 4.10	Data training dan data testing - LSTM tahunan.....	80
Gambar 4.11	Arsitektur LSTM Prediksi Tiga Tahunan.....	83
Gambar 4.12	Data training dan data testing - SVR tiga tahunan.....	87
Gambar 4.13	Data training dan data testing- LSTM tiga tahunan.....	87

INTISARI

Realisasi Investasi merupakan salah satu faktor yang mendukung pertumbuhan ekonomi. Peningkatan realisasi investasi memiliki dampak positif yang sangat luas antara lain mengurangi jumlah pengangguran, mengurangi jumlah penduduk miskin maupun peningkatan pendapatan negara. Untuk itu diperlukan suatu metode peramalan untuk melakukan prediksi nilai realisasi investasi yang dapat dijadikan referensi pengambilan keputusan dalam pembuatan kebijakan untuk meningkatkan capaian nilai realisasi investasi.

Pada penelitian ini akan menggunakan metode SVR dan LSTM dalam melakukan prediksi terhadap Nilai Realisasi Investasi. Data yang digunakan dari tahun 2007 sampai 2020. Prediksi akan dilakukan untuk jangka pendek bulanan, jangka menengah tahunan dan dan jangka panjang tiga tahunan. Hasil prediksi akan dibandingkan dengan nilai RMSE sehingga mendapatkan model dan hasil terbaik.

Hasil penelitian ini menunjukkan prediksi realisasi investasi bulanan menggunakan algoritma LSTM dengan optimasi Rmsprop memiliki hasil yang terbaik dibandingkan hasil prediksi tahunan dan tiga tahunan dengan menggunakan algoritma yang sama dan lebih baik disemua prediksi menggunakan algoritma SVR, dengan nilai RMSE 0,444.

Kata kunci: Realisasi Investasi, Support Vector Regression (SVR), Long Short Term Memory (LSTM), Grid Search, Regression

ABSTRACT

Investment realization is one of the factors that support economic growth. The increase in investment realization has a very broad positive impact, including reducing the number of unemployed, reducing the number of poor people and increasing state income. For this reason, a forecasting method is needed to predict the value of realized investment that can be used as a reference for decision making in policy making to increase the achievement of investment realization value.

In this study, the SVR and LSTM methods will be used to predict the Realized Value of Investment. The data used are from 2007 to 2020. Predictions will be made for the short term monthly, medium term annual and long term three years. The prediction results will be compared with the RMSE value so as to get the best model and results.

The results of this study indicate that the prediction of monthly investment realization using the LSTM algorithm with Rmsprop optimization has the best results compared to the annual and three-year predictions using the same algorithm and is better in all predictions using the SVR algorithm, with an RMSE value of 0.444.

Keyword: Investment Realization, Support Vector Regression (SVR), Long Short Term Memory (LSTM), Grid Search, Regression

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Penanaman Modal merupakan salah satu faktor untuk meningkatkan perekonomian Indonesia, karena dengan adanya penanaman modal dapat memperbanyak lapangan pekerjaan dan kesempatan kerja, peningkatan pendapatan, dan peningkatan kesejahteraan penduduk. Terciptanya barang modal dari kegiatan investasi akan menciptakan faktor produksi dan lapangan kerja baru sehingga siklus perekonomian akan berlangsung kearah yang positif (Prasetyo, 2009).

Untuk meningkatkan penanaman modal Pemerintah menerbitkan berbagai macam kebijakan percepatan berusaha. Salah satu kebijakan Pemerintah tersebut adalah ditetapkannya PP Nomor 24 Tahun 2018 tentang Pelayanan Berusaha Terintegrasi Secara Elektronik dan Perpres RI Nomor 91 Tahun 2017 Tentang Percepatan Pelaksana Berusaha dimana didalamnya mengatur pelaksanaan perizinan berusaha menggunakan Sistem Online Single Submission (OSS). Selain itu penetapan tata ruang yang clear and clean, ketersediaan tenaga kerja, ketersediaan Infrastruktur juga merupakan faktor yang mempengaruhi adanya penanaman modal. Sehingga kebijakan-kebijakan tersebut diharapkan akan menciptakan iklim investasi yang kondusif bagi penanam modal

BKPM dalam hal ini lembaga ditingkat pusat yang mempunyai urusan terkait penanaman modal mempunyai tanggung jawab terhadap capaian realisasi

penanaman modal. Selain menjadi tanggung jawab Pemerintah Pusat di tingkat Provinsi dan Kabupaten Kota mempunyai tanggung jawab yang sama sesuai dengan kewenangannya. Oleh karena itu dari tingkat Pusat dalam hal ini BKPM hingga Kabupaten Kota yang diampu oleh Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPSTP) dalam rencana kerjanya memiliki target terkait dengan capaian realisasi investasi.

BKPM dalam rencana strategis lima tahunan didalamnya terdapat indikator capaian realisasi investasi. Untuk tahun 2015-2019 terdapat 2 kali perubahan untuk beberapa penyesuaian salah satunya penyesuaian target realisasi investasi. Penentuan target tersebut sudah didasarkan pada potensi penanaman modal dan risiko ketidakpastian dan perlambatan ekonomi dunia. Target yang sudah ditetapkan akan dibandingkan dengan realisasi yang dihasilkan, dengan harapan realisasi akan melebihi target yang telah ditetapkan. Berdasarkan tabel 1.1 dapat kita lihat untuk realisasi investasi tahun 2014 hingga tahun 2019 sebagian besar melebihi target yang telah ditetapkan kecuali pada tahun 2018 realisasi investasi tercapai 94,3%.

Tabel 1.1 Perkembangan Penanaman Modal Tahun 2014 – 2019
(Peraturan BKPM RI Nomor 2 tahun 2020 tentang rencana strategis BKPM tahun 2020-2025), diolah

Keterangan	Tahun					
	2014	2015	2016	2017	2018	2019
Realisasi PMA (Rp Triliun)	307,0	365,9	396,6	430,5	392,7	423,1
Target PMA (Rp Triliun)	297,3	343,7	386,4	429	467,4	483,7
Realisasi PMDN (Rp Triliun)	156,1	179,5	216,2	262,3	328,6	386,5
Target PMDN (Rp Triliun)	159,3	175,8	208,4	249,8	297,6	308,3
Total Realisasi (Rp Triliun)	463,1	545,4	612,8	692,8	721,3	809,6
Total Target (Rp Triliun)	456,6	519,5	594,8	678,8	765	792
Capaian Realisasi (%)	101,4%	105%	103%	102,1%	94,3%	102,2%
Pertumbuhan (%)	16,2%	17,8%	12,4%	13,1%	4,1%	12,2%

Demikian halnya dengan provinsi dan kabupaten/Kota juga menetapkan target realisasi investasi pada rencana strategis untuk lima tahun kedepan. Rencana strategis 5 tahunan, dalam pelaksanaannya target yang ditetapkan seringkali hanya berdasarkan insting atau menambahkan dari target tahun sebelumnya sehingga bila dilihat dari capaian ada beberapa realisasi yang jauh melebihi target. Target dan realisasi investasi tahun 2018 sampai 2020 dapat dilihat pada tabel 1.2.

Tabel 1.2 Data perkembangan Penanaman Modal Tahun 2015 – 2020
(Sumber DPMPSTSP Kota Magelang, diolah)

Uraian Target	Tahun				
	2016	2017	2018	2019	2020
Pertumbuhan Nilai Investasi PMDN					
Target (Rp Juta)	533.964	603.612	673.260	690.671	812.555
Realisasi (Rp Juta)	672.248	781.939	1.250.172	1.405.381	1.870.535
Pertumbuhan investasi PMDN	44,78%	68,41%	169,25%	202,68%	302,86%
Target pertumbuhan investasi PMDN	15%	15%	15%	15%	15%
Pertumbuhan Nilai Investasi PMA					
Target (Rp Juta)	203.512	203.512	203.512	213.203	213.203
Realisasi (Rp Juta)	193.821	194.196	205.256	205.256	233.624
Pertumbuhan investasi PMA	0%	0,19%	5,9%	5,9%	20,54%
Target pertumbuhan investasi PMA	5%	5%	5%	10%	10%

Berdasarkan permasalahan realisasi investasi yang telah dijabarkan, dilakukan penelitian untuk melakukan prediksi realisasi investasi di Kota Magelang. Data capaian realisasi investasi sangat penting dalam rangka menggambarkan kondisi perekonomian suatu wilayah. (Edwin Setya Parahita, Junaedi, 2020). Penelitian ini akan mengidentifikasi dan menganalisis realisasi investasi secara berkala yang dibagi dalam beberapa segmentasi yaitu bulanan, tahunan dan tiga tahun. Data tersebut akan diproses menggunakan metode SVR dan LSTM. Identifikasi berguna

untuk mencari algoritma yang lebih baik untuk mengetahui prediksi realisasi investasi di masa yang akan datang. Penelitian ini diharapkan dapat memberikan kontribusi dalam proses menentukan target dan memberikan gambaran data realisasi investasi di Kota Magelang dalam kurun waktu jangka pendek, jangka menengah dan jangka panjang. Sehingga dapat mengambil langkah-langkah dalam upaya peningkatan realisasi investasi untuk meningkatkan pertumbuhan ekonomi.

Terdapat beberapa penelitian terkait realisasi penanaman modal (Edwin Setya Parahita, Junaedi, 2020) mengenai hal-hal yang mempengaruhi validitas data capaian realisasi investasi antara lain data investasi yang bersumber dari izin usaha yang kurang validasinya, tidak ada validasi lapangan dan adanya duplikasi data. (Ivonne Margi Immanuela, 2016) melakukan penelitian tentang faktor yang mempengaruhi kepercayaan pada layanan LKPM dengan menggunakan data kuesioner yang diisi oleh 389 responden. Data kuesioner dianalisis dengan menggunakan partial least squares-structural equation modeling (PLS-SEM) dengan hasil bahwa faktor yang mempengaruhi kepercayaan terhadap layanan LKPM adalah kualitas layanan LKPM online, keterbiasaan menggunakan layanan LKPM online, kepercayaan terhadap Pemerintah (BKPM). Demikian juga riset (Setyowati dan Fatimah, 2007) yang menunjukkan bahwa dalam jangka pendek capaian investasi dalam negeri tahun sebelumnya sangat berpengaruh terhadap capaian investasi dalam negeri tahun berikutnya. Iklim investasi yang dilanjutkan dengan capaian realisasi investasi berjalan dengan kondusif dan baik, maka dapat dipastikan hal tersebut dapat meningkatkan pertumbuhan ekonomi (Maharani, 2016)

Namun untuk penelitian mengenai investasi penanaman modal menggunakan data mining masih sangat sedikit. Penelitian di Indonesia yang menggunakan algoritma SVR dan LSTM untuk memprediksi realisasi investasi belum ada. Beberapa contoh penelitian Investasi Penanaman Modal di Indonesia yang menggunakan algoritma untuk analisis adalah sebagai berikut :

1. Metode K-Means Clustering untuk pengalokasian Penanam Modal Asing pada Badan Penanaman Modal Daerah Provinsi Jawa Tengah (Anis Wilasih, 2015)
2. Sistem Penentuan Tingkat Investasi menggunakan algoritma C4.5 (Yuli Murdianingsih, 2015)

Oleh karena itu penulis akan melihat penelitian-penelitian terkait data mining antara lain :

Penelitian dilakukan oleh M Abdul Dwiyanto Suyudi, Esmeralda C. Djamal dan Asri Maspupah (2019) menggunakan algoritma Recurrent Neural Network untuk melakukan prediksi harga saham. Data saham dari tahun 2012 hingga 2018 terdiri dari 1218 data. Data dilakukan pra-prosesing untuk mengubah data yang belum terstruktur menjadi data yang terstruktur. tersebut dibagi menjadi dua bagian yaitu 80% untuk data latih dan 20% untuk data uji. Pengujian dilakukan dengan beberapa cara yaitu : model optimasi dengan membandingkan dua parameter yaitu pengujian jumlah data dan juga pengujian learning rate. Hasil dari penelitian ini menunjukkan bahwa prediksi harga saham menggunakan RNN mendapatkan hasil kurang baik. Menggunakan optimasi Adam dengan nilai learning rate 0.001 dan

epoch 200 memperoleh akurasi untuk data latih sebesar 94.16% dan data uji sebesar 55.26%.

Penelitian dilakukan oleh Ni Putu Ratindia Apriyanti, I Ketut Gede Darma Putra dan I Made Suwija Putra (2020) menggunakan algoritma Support Vector Regression. Data bersumber dari Badan Pusat Statistik (BPS) Provinsi Bali dan data yang digunakan per tahun dengan jumlah 13 data. Peramalan menggunakan Metode Support Vector Regression dengan menggunakan dua kernel yaitu Kernel Polynomial dan Kernel RBF. Kesimpulan dari hasil peramalan yang telah dilakukan dengan mengimplementasikan Metode Support Vector Regression menggunakan kernel Polynomial memiliki hasil akurasi yang lebih baik, yaitu dengan nilai MAPE sebesar 7.95% sedangkan kernel RBF menghasilkan nilai MAPE sebesar 13.35% dengan pernyataan "Bagus".

Penelitian dilakukan oleh Arynda Kusuma Dewi (2020) menggunakan algoritma Support Vector Regresion untuk melakukan prediksi laju pertumbuhan penduduk, Studi Kasus Kota Malang. Data diperoleh dari Dispendukcapil Kota Malang berupa data times series sejumlah 72 data. Pengujian menggunakan kernel linear dan kernel Gaussian RBF. Penelitian ini menghasilkan nilai MAPE dengan menggunakan kernel linear sebesar 0,0985% Gaussian RBF sebesar 0,38192%.

Penelitian dilakukan oleh Satyam Gangwar, Vikram Bali dan Ajay Kumar (2019) menggunakan algoritma LSTM dan SVM untuk peramalan kecepatan angin. Sample penelitian dari [Weather archive Jena | Kaggle](#). Pada preprocessing data yang digunakan hanya lima parameter. Kedua metode menggunakan data set yang sama.

Hasil penelitian RMSE untuk LSTM adalah 0,427 dan nilai RMSE untuk SVM adalah 0,768. Algoritma LSTM memiliki hasil yang lebih baik daripada SVM.

Penelitian dilakukan oleh Muhammad Wildan Putra Aldi, Jondri, Annisa Aditsania (2018) menggunakan algoritma LSTM untuk melakukan prediksi harga bitcoin. Data diambil dari blockchain.info sebanyak 728 yang terbagi dalam 2 kombinasi data yaitu 50% untuk data latih dan data uji serta 70% data latih dan 30% data uji. Preprocessing teknik normalisasi dengan *mix-max scaling*, inialisasi parameter terdiri dari komposisi data, jumlah pola *time series*, jumlah neuron hidden dan epoch. Pada pengujian tersebut didapatkan hasil yang terbaik yaitu dengan komposisi data latih 70% dan data uji 30%, parameter 1 pola *time series*, jumlah 25 neuron hidden, dan *max epoch* adalah 100 dengan akurasi rata-rata pada data latih 95,36% dan data testing 93,5%.

Penelitian dilakukan oleh Jannatun Khustia Lubis, Iqbal Kharisudin (2021) menggunakan algoritma *Long Short Term Memory* dan *Generalized Autoregressive Conditional Heteroscedasticity* untuk pemodelan data saham. Data diambil dari data harga penutupan saham perusahaan Alphabet Inc mulai tanggal 01 Juli 2015- 08 September 2020. Penelitian ini menerapkan pembelajaran *supervised learning*, kemudian dilakukan normalisasi data dengan rumus *min max* dan untuk pembagian data *training* dan *testing* adalah 80% dibanding 20%. Arsitektur jaringan menggunakan *input layer*, *LSTM layer* dan *output layer* dengan parameter sejumlah epoch dan sejumlah *batch size*. Dilakukan optimasi dengan Adam, Adamax, dan RMSProp serta sejumlah hidden neuron dan performance diukur dengan RMSE.

Penelitian yang dilakukan oleh Soffa Zahara, Sugianto, M. Bahril Ilmiddafiq yang melakukan Prediksi Indeks Harga Konsumen Menggunakan Metode Long Short Term Memory (LSTM) Berbasis Cloud Computing. Data diambil dari website Dinas Perdagangan dan Perindustrian Provinsi Jawa Timur yaitu <http://siskaperbapo.com/> dengan periode pengambilan tahun 2014 sampai 2018. Preprocessing dilakukan dengan pemrosesan data yang hilang kemudian tahap selanjutnya yaitu scaling. Untuk pembagian data training dan testing 70% dibanding 30%. Dilakukan pengujian menggunakan 7 algoritma optimasi yang akan dibandingkan kinerjanya. Nesterov Adam (Adam) mempunyai tingkat akurasi paling tinggi dibanding dengan algoritma lainnya dengan nilai RMSE terkecil yaitu 4.088.

Penelitian yang akan dilakukan Adhib Arfan, Lussiana ETP melakukan prediksi harga saham menggunakan algoritma LSTM dengan SVR. Penelitian ini menggunakan 2 data yaitu data saham dan harga kurs. Data saham yang digunakan diambil dari situs yahoo finance, sedangkan harga kurs yang digunakan diambil dari situs OFX. Proses normalisasi yang digunakan adalah min-max scalling pada proses perancangan menggunakan beberapa parameter antara lain : hidden layer, neuron, epoch, dan batch size. Saat pelatihan nilai loss dioptimasi dengan menggunakan RMSProp. Dari hasil pengujian LSTM mampu menanggulangi ketergantungan jangka panjang dan mampu memprediksi harga saham dengan hasil yang akurat.

Penelitian yang dilakukan oleh Drajat Indra Purnama melakukan peramalan jumlah penumpang dengan SVR. Data yang digunakan adalah data penumpang setiap bulan dari periode Januari 2006 sampai dengan November 2019. Sebelum

dilakukan penelitian dilakukan uji linearitas menggunakan uji Ramsey RESET yang kemudian menentukan lag berpengaruh menggunakan plot PACF. Sedangkan untuk model SVR sendiri menggunakan kernel RBF yang penentuan parameternya menggunakan metode grid search dan ketepatan peramalan diuji dengan MAPE. Hasil pemodelan menghasilkan tingkat akurasi yang baik dengan nilai MAPE data training sebesar 5,01 persen dan nilai MAPE data testing sebesar 16,65 persen.

Penelitian yang dilakukan oleh Aulia Asyiva yang melakukan prediksi laju inflasi dengan menggunakan metode SVR dengan kernel RBF. Menggunakan data indeks harga konsumen tahun 2010-2018 yang diambil dari website Bank Indonesia. Pembagian data set 72 data training dan 36 data set, yang kemudian dilakukan normalisasi dengan metode min-max. Pemodelan SVR menggunakan kernel RBF dimana penentuan parameternya menggunakan grid search optimization. Hasil pengujian menunjukkan nilai RMSE yang diperoleh sebesar 0,00000177 dengan hasil sangat baik.

Berdasarkan review penelitian terkait yang sudah dilakukan membuktikan bahwa metode SVR terbukti cocok digunakan untuk data nonlinear dan pada masalah regresi, sedangkan untuk mengatasi masalah yang muncul data nonlinear digunakan kernel dimana pada penelitian ini akan digunakan kernel Radial Basis Function dan kernel Polynomial. Selain itu pada metode SVR juga sering muncul masalah untuk menentukan parameter yang optimal maka untuk mengatasi hal tersebut akan digunakan metode grid search yang dibagi dalam dua tahap untuk mempersingkat waktu pengolahan yaitu dengan tahap loose grid dan finer grid.

Sedangkan Long Short Term Memory Neural Network (LSTM) merupakan salah satu jenis RNN. LSTM menyimpan informasi terhadap pola-pola pada data. LSTM dapat mempelajari data mana saja yang akan disimpan dan data mana saja yang akan dibuang, karena pada setiap neuron LSTM memiliki beberapa gates yang mengatur memori pada setiap neuron itu sendiri. LSTM banyak digunakan pada data time series. Untuk memperoleh hasil optimal akan dilakukan optimasi menggunakan algoritma Adam dan RMSProp.

Melihat kelebihan dari kedua metode tersebut maka akan dilakukan perbandingan terhadap kedua metode tersebut dalam melakukan prediksi terhadap realisasi investasi.

1.2. Rumusan Masalah

Dari latar belakang tersebut, rumusan masalah yang akan dibahas pada penelitian ini adalah :

1. Bagaimana Akurasi hasil prediksi nilai realisasi investasi menggunakan algoritma SVR dan Algoritma LSTM ?
2. Bagaimana perbandingan hasil prediksi nilai realisasi investasi menggunakan algoritma SVR dan LSTM untuk jangka pendek, menengah dan jangka panjang?
3. Algoritma manakah yang lebih unggul antara algoritma SVR dibandingkan dengan algoritma LSTM ?

1.3 Batasan Masalah

Batasan masalah dari penelitian adalah sebagai berikut :

1. Data investasi diambil dari data realisasi investasi pada DPMPSTP Kota Magelang
2. Data realisasi investasi yang digunakan hanya data realisasi investasi Penanaman Modal Dalam Negeri (PMDN) pada tahun 2007 - 2020.
3. Pada algoritma SVR akan menggunakan kernel RBF dan kernel Polynomial, sedangkan penentuan parameter menggunakan metode grid search.
4. Pada Algoritma LSTM dilakukan optimasi menggunakan algoritma Adam dan Rmsprop serta menggunakan parameter hidden layer, neuron, batch size dan epoch.
5. Prediksi dilakukan dalam 3 skenario yaitu bulanan, tahunan dan tiga tahunan. Skenario data time series diterapkan untuk prediksi tahunan dan tiga tahunan.
6. Data diolah menggunakan Microsoft Excel 2013 dan python.

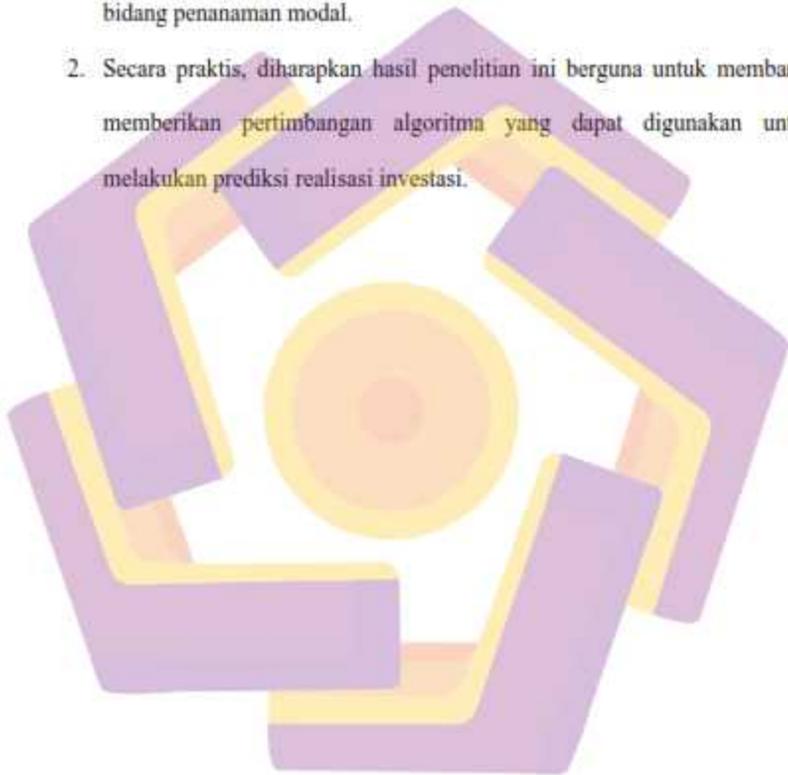
1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disebutkan sebelumnya, tujuan penelitian ini adalah mengetahui algoritma yang lebih sesuai dari algoritma SVR dan LSTM untuk melakukan prediksi nilai realisasi investasi.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah :

1. Secara teoritis, memberikan pengetahuan baru tentang prediksi nilai realisasi investasi dan memperkaya penelitian pengolahan data mining di bidang penanaman modal.
2. Secara praktis, diharapkan hasil penelitian ini berguna untuk membantu memberikan pertimbangan algoritma yang dapat digunakan untuk melakukan prediksi realisasi investasi.



BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Penelitian di bidang penanaman modal yang menggunakan algoritma data mining di Indonesia masih sangat terbatas sehingga penulis melakukan komparasi dengan penelitian dibidang lain terkait metode yang akan digunakan.

Penelitian dilakukan oleh M Abdul Dwiyanto Suyudi, Esmeralda C. Djamil dan Asri Maspupah (2019) menggunakan algoritma Recurrent Neural Network untuk melakukan prediksi harga saham. Data saham dari tahun 2012 hingga 2018 terdiri dari 1218 data. Data dilakukan pra-prosesing untuk mengubah data yang belum terstruktur menjadi data yang terstruktur, tersebut dibagi menjadi dua bagian yaitu 80% untuk data latih dan 20% untuk data uji. Pengujian dilakukan dengan beberapa cara yaitu : model optimasi dengan membandingkan dua parameter yaitu pengujian jumlah data dan juga pengujian learning rate. Hasil dari penelitian ini menunjukkan bahwa prediksi harga saham menggunakan RNN mendapatkan hasil kurang baik. Menggunakan optimasi Adam dengan nilai learning rate 0.001 dan epoch 200 memperoleh akurasi untuk data latih sebesar 94.16% dan data uji sebesar 55.26%. Kekurangan dari penelitian ini hanya menggunakan dua parameter saja yaitu learning rate dan epoch. Pada penelitian yang akan dilakukan akan menggunakan metode optimasi Adam dan parameter yang digunakan terdiri dari hidden layer, neuron, batch size dan epoch.

Penelitian dilakukan oleh Ni Putu Ratindia Apriyanti, I Ketut Gede Darma Putra dan I Made Suwija Putra (2020) menggunakan algoritma Support Vector Regression. Data bersumber dari Badan Pusat Statistik (BPS) Provinsi Bali dan data yang digunakan per tahun dengan jumlah 13 data. Peramalan menggunakan Metode Support Vector Regression dengan menggunakan dua kernel yaitu Kernel Polynomial dan Kernel RBF. Kesimpulan dari hasil peramalan yang telah dilakukan dengan mengimplementasikan Metode Support Vector Regression menggunakan kernel Polynomial memiliki hasil akurasi yang lebih baik, yaitu dengan nilai MAPE sebesar 7.95% sedangkan kernel RBF menghasilkan nilai MAPE sebesar 13.35% dengan pernyataan "Bagus". Kekurangan dari penelitian ini tidak dijelaskan proses penentuan parameter C yang digunakan pada kernel polynomial maupun kernel RBF yang ternyata sangat mempengaruhi hasil peramalan. Pada penelitian yang akan dilakukan memperbandingkan 2 algoritma yaitu LSTM dan SVR. Untuk metode SVR fokus kepada satu kernel yaitu kernel RBF karena merupakan kernel yang paling sering digunakan, dan dengan pemilihan nilai α yang tepat, kernel ini dapat memiliki akurasi yang sama dengan kernel linier dan kernel polynomial. Untuk Penentuan parameter dari kernel RBF ini akan menggunakan metode grid search agar mendapatkan nilai parameter yang optimal.

Penelitian dilakukan oleh Arynda Kusuma Dewi (2020) menggunakan algoritma Support Vector Regression untuk melakukan prediksi laju pertumbuhan penduduk, Studi Kasus Kota Malang. Data diperoleh dari Dispendukcapil Kota Malang berupa data times series sejumlah 72 data. Pengujian menggunakan kernel linear dan kernel Gaussian RBF. Penelitian ini menghasilkan nilai MAPE dengan

menggunakan kernel linear sebesar 0,0985% Gaussian RBF sebesar 0,38192%. Kekurangan pada penelitian ini penentuan parameter pada kernel berdasarkan penelitian sebelumnya. Untuk penelitian yang akan dilakukan penentuan parameter kernel menggunakan metode gridsearch.

Penelitian dilakukan oleh Satyam Gangwar, Vikram Bali dan Ajay Kumar (2019) menggunakan algoritma LSTM dan SVM untuk peramalan kecepatan angin. Sample penelitian dari [Weather archive Jena | Kaggle](#). Pada preprocessing data yang digunakan hanya lima parameter. Kedua metode menggunakan data set yang sama. Hasil penelitian RMSE untuk LSTM adalah 0,427 dan nilai RMSE untuk SVM adalah 0,768. Algoritma LSTM memiliki hasil yang lebih baik daripada SVM. Penelitian ini tidak menginformasikan jumlah data set yang telah dikumpulkan dan pembagian dataset tersebut untuk kebutuhan testing dan training. Pada Penelitian yang akan dilakukan akan menggunakan pembagian data set training dan testing sebanyak 70% dan 30%.

Penelitian dilakukan oleh Muhammad Wildan Putra Aldi, Jondri, Annisa Aditsania (2018) menggunakan algoritma LSTM untuk melakukan prediksi harga bitcoin. Data diambil dari [blockchain.info](#) sebanyak 728 yang terbagi dalam 2 kombinasi data yaitu 50% untuk data latih dan data uji serta 70% data latih dan 30% data uji. Preprocessing teknik normalisasi dengan mix-max scalling, inialisasi parameter terdiri dari komposisi data, jumlah pola time series, jumlah neuron hidden dan epoch. Pada pengujian tersebut didapatkan hasil yang terbaik yaitu dengan komposisi data latih 70% dan data uji 30%, parameter 1 pola time series, jumlah 25 neuron hidden, dan max epoch adalah 100 dengan akurasi rata-rata

pada data latih 95,36% dan data testing 93,5%. Pada penelitian yang akan dilakukan menggunakan kombinasi pembagian dataset 70% training dan 30% testing serta parameter yang digunakan hidden layer, neuron, epoch dan batch size.

Penelitian dilakukan oleh Jannatun Khustia Lubis, Iqbal Kharisudin (2021) menggunakan algoritma Long Short Term Memory dan Generalized Autoregressive Conditional Heteroscedasticity untuk pemodelan data saham. Data diambil dari data harga penutupan saham perusahaan Alphabet Inc mulai tanggal 01 Juli 2015- 08 September 2020. Penelitian ini menerapkan pembelajaran supervised learning, kemudian dilakukan normalisasi data dengan rumus min max dan untuk pembagian data training dan testing adalah 80% dibanding 20%. Arsitektur jaringan menggunakan input layer, LSTM layer dan output layer dengan parameter sejumlah epoch dan sejumlah batch size. Dilakukan optimasi dengan Adam, Adamax, dan RMSProp serta sejumlah hidden neuron dan performance diukur dengan RMSE. Kekurangan dari penelitian ini adalah yang pertama tidak diinformasikan hasil optimasi yang paling akurat dari 3 optimasi yang ada, kedua tidak diinformasikan metode yang paling akurat antara LSTM dan GARCH. Penelitian yang akan dilakukan menggunakan satu metode optimasi adam yang merupakan gabungan metode Adagrad dan RMSprop dan dari review penelitian sebelumnya memiliki akurasi yang baik. Selain itu untuk parameter yang digunakan ditambahkan hidden layer dan neuron.

Penelitian yang dilakukan oleh Soffa Zahara, Sugianto, M. Bahril Ilmidafiq yang melakukan Prediksi Indeks Harga Konsumen Menggunakan Metode Long Short Term Memory (LSTM) Berbasis Cloud Computing. Data diambil dari

website Dinas Perdagangan dan Perindustrian Provinsi Jawa Timur yaitu <http://siskaperbapo.com/> dengan periode pengambilan tahun 2014 sampai 2018. Preprocessing dilakukan dengan pemrosesan data yang hilang kemudian tahap selanjutnya yaitu scaling. Untuk pembagian data training dan testing 70% dibanding 30%. Dilakukan pengujian menggunakan 7 algoritma optimasi yang akan dibandingkan kinerjanya. Nesterov Adam (Adam) mempunyai tingkat akurasi paling tinggi dibanding dengan algoritma lainnya dengan nilai RMSE terkecil yaitu 4.088. Pada penelitian yang akan dilakukan akan mengukur akurasi metode SVR dan LSTM dalam melakukan prediksi realisasi investasi dan fokus menggunakan satu metode optimasi Adam yang pada review penelitian tersebut memiliki tingkat akurasi yang paling tinggi.

Penelitian yang akan dilakukan Adhib Arfan, Lussiana ETP melakukan prediksi harga saham menggunakan algoritma LSTM dengan SVR. Penelitian ini menggunakan 2 data yaitu data saham dan harga kurs. Data saham yang digunakan diambil dari situs yahoo finance, sedangkan harga kurs yang digunakan diambil dari situs OFX. Proses normalisasi yang digunakan adalah min-max scalling pada proses perancangan menggunakan beberapa parameter antara lain : hidden layer, neuron, epoch, dan batch size. Saat pelatihan nilai loss dioptimasi dengan menggunakan RMSProp. Dari hasil pengujian LSTM mampu menanggulangi ketergantungan jangka panjang dan mampu memprediksi harga saham dengan hasil yang akurat. Kekurangan dari penelitian ini tidak disebutkan proses pada metode SVR. Pada penelitian yang akan dilakukan menggunakan dua metode dengan SVR dan LSTM. Pada metode SVR penentuan parameter akan menggunakan grid search agar

mendapatkan hasil yang optimal sedangkan untuk metode LSTM akan dilakukan optimasi menggunakan metode Adam yang mengambil kelebihan dari metode RMSProp dan Adagrad.

Penelitian yang dilakukan oleh Drajat Indra Purnama melakukan peramalan jumlah penumpang dengan SVR. Data yang digunakan adalah data penumpang setiap bulan dari periode Januari 2006 sampai dengan November 2019, dengan pembagian data training dan data testing adalah 90% dan 10%. Sebelum dilakukan penelitian dilakukan uji linearitas menggunakan uji Ramsey RESET yang kemudian menentukan lag berpengaruh menggunakan plot PACF. Sedangkan untuk model SVR sendiri menggunakan kernel RBF yang penentuan parameternya menggunakan metode grid search dan ketepatan peramalan diuji dengan MAPE. Hasil pemodelan menghasilkan tingkat akurasi yang baik dengan nilai MAPE data training sebesar 5,01 persen dan nilai MAPE data testing sebesar 16,65 persen. Penelitian yang akan dilakukan menggunakan pembagian data training dan testing sebesar 70% dan 30% dengan menggunakan dua metode yaitu SVR dan LSTM yang akan dibandingkan hasilnya mana yang memiliki performa yang terbaik. Untuk model SVR menggunakan kernel RBF dan metode search grid untuk menentukan parameternya. Pemilihan metode search grid yang dibagi menjadi 2 tahap yaitu loose grid dan finer grid karena pada penelitian yang direview memiliki hasil yang bagus.

Penelitian yang dilakukan oleh Aulia Asyiva yang melakukan prediksi laju inflasi dengan menggunakan metode SVR dengan kernel RBF. Menggunakan data indeks harga konsumen tahun 2010-2018 yang diambil dari website Bank

Indonesia. Pembagian data set 72 data training dan 36 data set, yang kemudian dilakukan normalisasi dengan metode min-max. Pemodelan SVR menggunakan kernel RBF dimana penentuan parameternya menggunakan grid search optimization. Hasil pengujian menunjukkan nilai RMSE yang diperoleh sebesar 0,00000177 dengan hasil sangat baik. Penelitian yang akan dilakukan menggunakan 2 metode yaitu SVR dan LSTM untuk diperbandingkan akurasi. Metode SVR sendiri dalam penentuan parameter menggunakan metode grid search yang dibagi dengan dua tahap yaitu loose grid dan finer grid untuk mengurangi lama waktu proses.

Penelitian yang dilakukan oleh Ekky Novriza Alam yang melakukan prediksi kebutuhan Obat dengan menggunakan algoritma Support Vector Regression (SVR) dan Long Short Term Memory (LSTM). Data yang digunakan diambil dari Rumah Sakit Ibu dan Anak (RSIA) Zainab Riau. Proses Preprocessing data dilakukan dengan mengisi data kosong, menghilangkan duplikasi dan menghapus data outlier, sedangkan transformasi data dilakukan dengan menggunakan normalisasi min max dan differencing untuk data yang tidak stasioner. Model prediksi dibagi menjadi beberapa skenario antara lain prediksi harian dan mingguan pada prediksi mingguan menggunakan tujuh hari sebelumnya sebagai input dan tujuh hari berikutnya sebagai output. Hasil penelitian menunjukkan algoritma LSTM memiliki nilai RMSE yang lebih baik pada prediksi mingguan pada jenis obat tertentu. Penelitian yang akan dilakukan menggunakan dua algoritma yaitu SVR dan LSTM dan pada SVR penentuan parameternya menggunakan metode grid search sedangkan parameter LSTM sudah ditentukan

dan membandingkan dua macam optimasi yaitu optimasi adam dan optimasi Rmsprop. Skenario prediksi meliputi skenario harian, tahunan dan tiga tahunan.

Penelitian yang dilakukan oleh Ahmad Ashril Rizal dan Siti Soraya yang melakukan multi time step prediksi dengan menggunakan RNN LSTM. Data yang digunakan adalah data kunjungan wisata pada tahun 2009 sampai 2014. Pembagian data set 70% data training dan 30% data testing. Normalisasi data pada interval $(-1,1)$ yang merupakan batas nilai untuk fungsi aktivasi hyperbolic tangent. Training RNN LSTM dengan menggunakan tiga model memberikan hasil yang beragam. Dari ketiga model LSTM yang dilakukan yaitu LSTM regression, LSTM dengan sliding window dan LSTM dengan time steps. Pada penelitian ini akan menggunakan algoritma SVR dan LSTM dan proses training akan dilakukan dengan time step.

2.2 Landasan Teori

2.2.1 Investasi

Penanaman modal merupakan istilah-istilah yang dikenal, baik dalam kegiatan bisnis sehari-hari maupun dalam bahasa perundang-undangan. Istilah investasi merupakan istilah yang lebih populer dalam dunia usaha, sedangkan istilah penanaman modal lebih banyak digunakan dalam bahasa perundang-undangan. Investasi berasal dari kata invest yang berarti menanam, menginvestasikan atau menanam uang (Andreas Halim, 2003). Penanaman Modal Dalam Undang-Undang No. 25 Tahun 2007 tentang Penanaman Modal menyebutkan bahwa "Penanaman modal adalah segala bentuk kegiatan menanam

modal, baik oleh penanam modal dalam negeri maupun penanam modal asing untuk melakukan usaha di wilayah negara Republik Indonesia.”

Ada 2 bentuk yaitu Penanaman Modal yaitu :

1. Penanaman Modal Dalam Negeri (PMDN)

Pengertian PMDN menurut Pasal 1 angka 2 Undang-Undang Penanaman Modal adalah: “kegiatan menanam modal untuk melakukan usaha di wilayah negara Republik Indonesia yang dilakukan oleh penanam modal dalam negeri dengan menggunakan modal dalam negeri.”

2. Penanaman Modal asing (PMA)

Pengertian penanaman modal asing dapat ditemukan pada pasal 1 angka 3 Undang-Undang Penanaman Modal. Penanaman modal asing ialah: “kegiatan menanam modal untuk melakukan usaha di wilayah negara Republik Indonesia yang dilakukan oleh penanam modal asing, baik yang menggunakan modal asing sepenuhnya maupun yang berpatungan dengan penanam modal dalam negeri”.

2.2.2 Outlier

Outlier merupakan data yang memiliki karakteristik unik yang sangat berbeda jauh dari data lainnya dan muncul dalam bentuk nilai ekstrim baik untuk sebuah variabel tunggal atau variabel kombinasi (Imam, 2011). Sebab munculnya data outlier dapat disebabkan oleh beberapa hal antara lain : karena kesalahan analisis atau prosedur dalam memasukkan data, kegagalan melakukan spesifikasi missing value ke dalam program komputer, mengambil sample data yang salah dan

sample yang diambil benar namun memiliki nilai yang ekstrim dan tidak berdistribusi normal (Ira, 2013). Identifikasi terhadap dapat dilakukan dengan beberapa cara antara lain dengan menggunakan boxplot. Metode box plot banyak digunakan dengan menggunakan nilai kuartil dan nilai jangkauan. Kuartil dibagi menjadi kuartil 1, 2 dan 3 yang akan membagi sebuah urutan data menjadi empat bagian. Sedangkan Jangkauan (IQR) merupakan interquartile range yang merupakan selisih antara kuartil 1 dan kuartil 3, atau dapat ditulis $IQR = Q3 - Q1$.

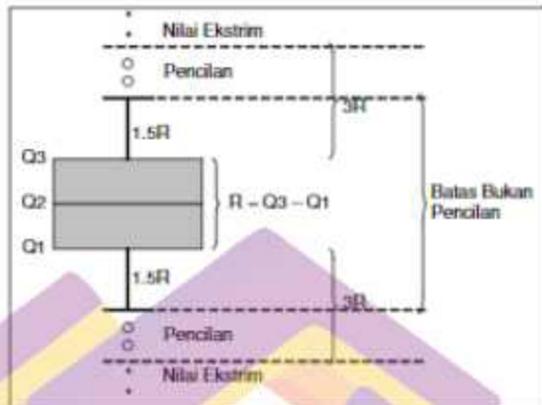
Outlier dapat dibedakan menjadi dua yaitu (Gusti, 2014)

1. Minor outlier

Minor outlier merupakan outlier yang berada pada range antara kuartil pertama dikurangi 3 kali jangkauan kuartil dan kuartil pertama dikurangi 1.5 kali jangkauan kuartil. Range tersebut untuk outlier bawah, sedangkan untuk outlier atas berada pada range kuartil 3 ditambah 1.5 jangkauan kuartil dan kuartil 3 ditambah 3 kali jangkauan kuartil.

2. Mayor outlier

Mayor outlier merupakan outlier yang nilainya lebih kecil dari nilai kuartil pertama dikurangi 3 kali jangkauan kuartil, dan lebih besar dari nilai kuartil ketiga ditambah 3 kali jangkauan antar kuartil.



Gambar 2.1 Identifikasi Pencilan Menggunakan IQR

Untuk menangani data outlier dapat dilakukan dengan beberapa cara antara lain :

1. Menghapus nilai outlier
2. Mengganti nilai outlier dengan nilai upper kuartil
3. Mengganti nilai outlier dengan nilai mean, median atau mode

2.2.3 Support Vektor Regression (SVR)

SVR merupakan penerapan SVM yang digunakan untuk kasus regresi yang outputnya berupa bilangan riil atau kontinu. Ide dasar dari SVR yaitu dengan menentukan set data yang dibagi menjadi data training dan data testing. Kemudian dari data training tersebut ditentukan suatu fungsi regresi dengan batasan deviasi tertentu sehingga dapat menghasilkan prediksi yang mendekati nilai aktual.

Fungsi regresi dari metode SVR linear adalah sebagai berikut :

$$f(x_i) = w \cdot x_i + b \quad (2.1)$$

Untuk mendapatkan generalisasi fungsi regresi $f(x)$ yang baik dapat dilakukan dengan meminimalkan nilai w melalui penyelesaian optimasi sebagai berikut :

$$\frac{1}{2} \|w\|^2 \quad (2.2)$$

dengan syarat : $y_i - (w \cdot x) - b \leq \epsilon$ dan $(w \cdot x) - b + y_i \leq \epsilon$

Pada persamaan (2) tersebut diasumsikan semua titik ada pada rentang nilai $\pm \epsilon$. Namun seringkali ada nilai-nilai kesalahan yang boleh melebihi nilai ϵ , karena adanya titik-titik yang dimungkinkan keluar dari nilai $\pm \epsilon$. Sehingga dibutuhkan soft margin yang biasa disebut variable slack (ξ_i, ξ_i^*) sehingga persamaannya menjadi :

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (2.3)$$

dengan syarat : $y_i - (w \cdot x) - b \leq \epsilon + \xi_i$ dan $(w \cdot x) - b + y_i \leq \epsilon + \xi_i^*$, ξ_i dan $\xi_i^* \geq 0$

Dengan menggunakan langrange multipliers serta kondisi Karush Kuhn Tucker pada persamaan (3) maka dihasilkan persamaan sebagai berikut :

$$L(\alpha, \alpha^*) = -\epsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i) + \sum_{i=1}^n (\alpha_i^* - \alpha_i) q_i - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) K(x_i, x_j) \quad (2.4)$$

Dengan syarat : $\sum_{i=1}^n (\alpha_i^* - \alpha_i) = 0$

$$0 \leq \alpha_i \leq C \text{ dan } 0 \leq \alpha_i^* \leq C$$

Pada umumnya data dalam dunia nyata kebanyakan bersifat nonlinear. SVR untuk kasus nonlinear memberikan pendekatan alternatif dengan cara melakukan pemetaan data x dari input space ke feature space dengan dimensi yang lebih tinggi melalui suatu fungsi φ sehingga $\varphi: \mathbf{x} \mapsto \varphi(\mathbf{x})$ (Santoso B, 2007). Sehingga fungsi regresi SVR nonlinear dituliskan sebagai berikut :

$$f(\mathbf{x}_i) = \sum(\alpha_i - \alpha_i^*) \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}) + b \quad (2.5)$$

Akan tetapi, kesulitan dalam pemetaannya adalah bahwa transformasi φ pada umumnya tidak diketahui dan sulit dipahami. Menurut (Vapnik, 1995), masalah ini dapat diatasi dengan kernel trick yaitu perkalian skalar (dot product) $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x})$ dalam feature space dapat digantikan fungsi kernel

$$K(\mathbf{x}_i, \mathbf{x}) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}) \quad (2.6)$$

dengan fungsi kernel tersebut mampu mendefinisikan secara implisit transformasi φ . Sehingga fungsi regresi dituliskan menjadi

$$f(\mathbf{x}_i) = \sum(\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (2.7)$$

Jenis fungsi kernel mempengaruhi parameter kernel SVR. Fungsi kernel harus diatur dengan benar karena dapat mempengaruhi akurasi regresi. Dalam penelitian ini, kernel yang digunakan pada model SVR adalah kernel Radial Basis Function (RBF) dan kernel Polynomial. Persamaan untuk kernel RBF sebagai berikut :

$$K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2) \quad (2.8)$$

SVR dengan fungsi kernel RBF memiliki dua parameter utama yang harus ditentukan yaitu C (cost), γ (gamma). Fungsi regresi dengan menggunakan kernel RBF adalah sebagai berikut :

$$f(\mathbf{x}_i) = \sum(\alpha_i - \alpha_i^*) \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2) + b \quad (2.9)$$

Sedangkan persamaan untuk kernel polynomial adalah :

$$K(\mathbf{x}_i, \mathbf{x}) = ((\mathbf{x}_i \cdot \mathbf{x}) + 1)^d \quad (2.10)$$

Pada kernel Polynomial parameter yang paling mempengaruhi adalah besaran degree. Fungsi regresi dengan menggunakan kernel polynomial adalah sebagai berikut :

$$f(\mathbf{x}_i) = \sum(\alpha_i - \alpha_i^*) ((\mathbf{x}_i - \mathbf{x}) + 1)^d + b \quad (2.11)$$

Metode yang digunakan untuk mencari parameter optimal pada penelitian ini menggunakan metode grid search.

Langkah-Langkah Support Vector Regression

Langkah-langkah perhitungan pada metode SVR adalah : (Arynda, 2020)

1. Menentukan jumlah iterasi maksimum, nilai lamda (λ), epsilon (ϵ), kompleksitas (C), gamma (γ) dengan menggunakan metode grid search dan untuk nilai lamda (λ) ditentukan secara langsung dan $\alpha_1 = 0$ dan $\alpha_1^* = 0$.
2. Menghitung nilai matriks hessian selama $i, j = 1, 2, \dots, n$ menggunakan persamaan 2.12 :

$$[R]_{ij} = K(x_i, x_j) + \lambda^2 \quad (2.12)$$

Keterangan :

$[R]_{ij}$ = Matrik Hessian

$K(x_i, x_j)$ = Fungsi kernel

λ = Lamda

i = Data ke i

j = Data ke j

4. Menghitung sequential learning dengan langkah dari a sampai c yaitu :
 - a. Menghitung nilai error dengan persamaan 2.13

$$E_i = y_i - \sum_{l=1}^n (\alpha_l - \alpha_l^*) [R]_{il} \quad (2.13)$$

- b. Menghitung perubahan nilai langrange multiplier dengan persamaan 2.14 dan 2.15

$$\delta \alpha_l^* = \min (\max[\gamma(E_l - \epsilon), -\alpha_l^*], C - \alpha_l^*) \quad (2.14)$$

$$\delta \alpha_i = \min (\max [\gamma(-E_i - \epsilon), -\alpha_i], C - \alpha_i) \quad (2.15)$$

- c. Menghitung update nilai langrange multiplier dengan persamaan 2.16.dan 2.17

$$\alpha_i^* = \delta \alpha_i^* + \alpha_i^* \quad (2.16)$$

$$\alpha_i = \delta \alpha_i + \alpha_i \quad (2.17)$$

Perhitungan dari langkah a sampai c akan berhenti bila telah memenuhi syarat stop condition yaitu mencapai iterasi maksimum atau $\max (|\delta \alpha_i^*|) < \epsilon$ dan $\max (|\delta \alpha_i|) < \epsilon$.

4. Menghitung fungsi regresi dengan menggunakan persamaan 2.18

$$f(x) = \sum_{i=1}^n (\alpha_i^* - \alpha_i) (K(x_i, x_j) + \lambda^2) \quad (2.18)$$

Keterangan :

α_i^*, α_i = Nilai Langrange multiplier pada iterasi terakhir yang memenuhi stop kondisi

$K(x_i, x_j)$ = Fungsi Kernel

λ = Lamda

$f(x)$ = Fungsi regresi

2.2.4 Grid Search

Grid search merupakan kombinasi parameter yang diujikan pada suatu model SVM untuk mencari nilai error dalam klasifikasi (Hsu, C. W, Chang, C.C., and Lin,CJ, 2016) Tujuannya adalah mengidentifikasi parameter optimal dalam data training, sehingga model tersebut mampu secara akurat memprediksi data

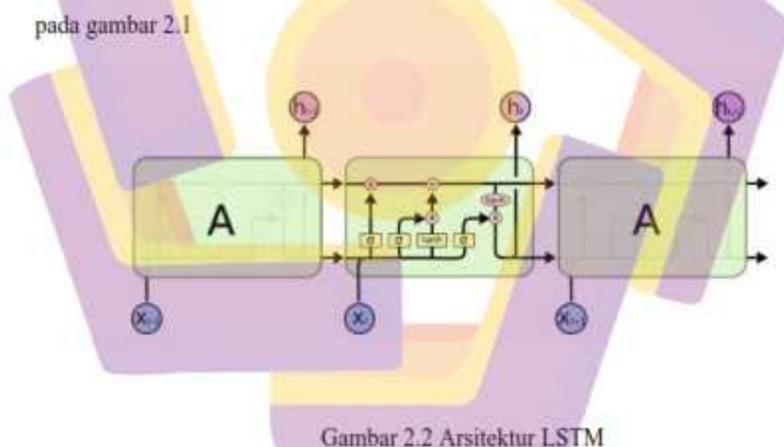
testing. Salah satu pendekatan dalam grid search adalah prosedur cross validation. Prosedur cross validation adalah membagi secara acak data training menjadi n subset yang berukuran sama. Satu subset digunakan sebagai data testing dan $n-1$ subset digunakan sebagai data training (Santoso B, 2007). Proses untuk melakukan grid search secara lengkap memerlukan waktu yang sangat lama sehingga Hsu (Hsu, C. W, Chang, C.C., and Lin.CJ, 2016), menyarankan untuk melakukan grid search dengan dua tahap yaitu loose grid dan finer grid. Loose grid adalah tahapan dimana pemilihan nilai C dan γ dengan pangkat bilangan bulat, sedangkan finer grid adalah tahapan selanjutnya dari loose grid dimana saat didapat nilai C dan γ dengan error terendah, maka finer grid menggunakan nilai di sekitaran C dan γ tersebut. Sejalan dengan hal tersebut, pada penelitian ini akan digunakan grid search dengan dua tahapan yaitu loose grid dan finer grid untuk menemukan nilai parameter optimal.

2.2.5 Long Short Term Memory (LSTM)

Recurrent Neural Network adalah bagian dari neural network yang memproses sejumlah data yang bersambung (sequential data). Secara umum, manusia tidak membuat keputusan dari awal setiap saat. Manusia selalu memperhitungkan informasi masa lalu dalam membuat keputusan. Seperti halnya manusia, Recurrent Neural Network juga memperhitungkan informasi masa lalu dalam membuat keputusannya.

LSTM merupakan sebuah evolusi dari arsitektur Recurrent Neural Network (RNN), di mana pertama kali diperkenalkan oleh Hochreiter & Schmidhuber (1997)

(Felix Gers, 2001) Hingga penelitian ini dilakukan banyak para peneliti yang terus mengembangkan arsitektur LSTM di berbagai bidang seperti dalam bidang speech recognition dan forecasting (Hendri H, 2014). RNN memiliki memori jangka pendek, yaitu tidak dapat membawa informasi yang diperoleh sebelumnya ke proses selanjutnya. Solusi dari keterbatasan yang dimiliki RNN adalah dengan menggunakan LSTM. LSTM merupakan special unit RNN yang dibuat untuk menanggulangi ketergantungan jangka panjang karena memiliki mekanisme internal yang disebut cell states dan gates yang dapat mengatur memori dalam setiap masukannya (Murtaza, 2017). Terdapat empat gates yaitu forget gate, input gate, cell gate dan output gate. Berikut adalah arsitektur dari LSTM yang dapat dilihat pada gambar 2.1



Gambar 2.2 Arsitektur LSTM

Alur kerja dari arsitektur LSTM pada gambar 2.1 dimulai dengan menentukan nilai yang tidak digunakan (forget gate), hal tersebut diakrenakan pada gate ini menggunakan sigmoid layer yang memiliki output dengan nilai antara 0 dan 1. Forget gate dihitung dengan menggunakan data output sebelumnya h_{t-1} dan

data input X_t saat ini. Persamaan pada forget dengan rumus sebagai berikut (Wildan, 2018) :

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.19)$$

Dimana:

f_t = forget gate

σ = fungsi sigmoid

W_f = nilai weight untuk forget gate

h_{t-1} = nilai output sebelum orde ke t

x_t = nilai input pada orde ke t

b_f = nilai bias pada forget gate

Sedangkan untuk nilai weight ditentukan dengan persamaan berikut :

$$W = \left(-\frac{1}{\sqrt{d}} \quad \frac{1}{\sqrt{d}} \right) \quad (2.20)$$

Dimana :

W = weight

d = jumlah variabel

Selanjutnya proses menentukan data masukan (input gates) yang terdiri dari dua gates dengan menggunakan fungsi aktivasi sigmoid dan fungsi aktivasi tanh, dengan rumus sebagai berikut:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.21)$$

Dimana:

i_t = input gate

σ = fungsi sigmoid

W_i = nilai weight untuk input gate

h_{t-1} = nilai output sebelum orde ke t

x_t = nilai input pada orde ke t

b_i = nilai bias pada input gate

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2.22)$$

Dimana:

\tilde{C}_t = nilai baru yang dapat ditambahkan ke cell state

\tanh = fungsi tanh

W_c = nilai weight untuk cell state

h_{t-1} = nilai output sebelum orde ke t

x_t = nilai input pada orde ke t

b_c = nilai bias pada cell state

Selanjutnya proses memperbaharui data masukan (cell gates) dengan menggabungkan nilai pada forget gate dan input gate, dengan rumus sebagai berikut:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (2.23)$$

Dimana:

C_t = cell state

f_t = forget gate

C_{t-1} = cell state sebelum orde ke t

i_t = input gate

\tilde{C}_t = nilai baru yang dapat ditambahkan ke cell state

Selanjutnya proses menentukan nilai keluaran (output gates). Terdapat dua proses pada proses menentukan nilai keluaran sama seperti dengan penentuan input gates, disini juga menggunakan fungsi aktivasi sigmoid yang memutuskan bagian mana dari cell state yang akan dihasilkan menjadi output dan aktivasi tanh untuk menempatkan nilai cell state pada range -1 sampai 1, dengan rumus sebagai berikut:

$$ot = \sigma(Wo.[ht-1, xt] + bo) \quad (2.24)$$

Dimana:

ot = output gate

σ = fungsi sigmoid

Wo = nilai weight untuk output gate

$ht-1$ = nilai output sebelum orde ke t

xt = nilai input pada orde ke t

bo = nilai bias pada output gate

$$ht = ot * \tanh(Ct) \quad (2.25)$$

Dimana:

ht = nilai output orde t

ot = output gate

\tanh = fungsi tanh

Ct = cell state

2.2.6 Normalisasi

Normalisasi merupakan transformasi data yang dilakukan untuk membuat nilai ada pada range tertentu sehingga tidak akan dijumpai data dengan nilai selisih yang

terlalu jauh (Abdul, 2019). Pada banyak penelitian transformasi data tersebut banyak menggunakan metode min-max normalization (Jannatun, 2021) dan pada penelitian ini juga akan menggunakan normalisasi minmax dengan range (-1,1) (Ahmad, 2018). Perhitungan normalisasi dengan menggunakan persamaan 2.25

$$x_i' = \frac{x_i - \min}{\max - \min} (\max R - \min R) + \min R \quad (2.25)$$

Keterangan :

x_i' = nilai hasil normalisasi

x_i = nilai awal sebelum normalisasi

\min = nilai terkecil dari atribut data

\max = nilai terbesar dari atribut data

$\min R$ dan $\max R$ = Nilai rentang data terkecil dan terbesar

2.2.7 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) digunakan untuk mengukur akurasi hasil prediksi dimana ukuran kesalahan diambil dari jumlah kuadrat kesalahan. Karena RMSE menggunakan jumlah kuadrat maka error yang besar akan menjadi semakin besar sehingga hasil prediksi yang akurat menurut RMSE adalah nilai yang semakin kecil atau mendekati nol. Nilai RMSE dapat dihitung dengan persamaan sebagai berikut

:

$$RMSE = \sqrt{\frac{\sum (y - \hat{y}_i)^2}{N}} \quad (15)$$

Dimana N merupakan jumlah data pada data testing. Sedangkan y merupakan nilai data aktual dan \hat{y}_i merupakan data hasil prediksi.

2.3. Keaslian Penelitian

Tabel 2.1 Matriks literatur review dan posisi penelitian
Prediksi Realisasi Investasi menggunakan metode Support Vector Regression dan Long Short Term Memory

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Prediksi Harga Saham menggunakan Metode Recurrent Neural Network	M Abdul Dwiyanto Suyudi, Esmeralda C. Djamal, Asri Maspupah, Seminar Nasional Aplikasi Teknologi Informasi (SNATI), Agustus 2019	Penelitian ini membangun sistem yang dapat memprediksi harga saham menggunakan faktor teknikal, dengan mengeluarkan perkiraan prediksi pada pergerakan saham beli atau jual	Hasil dari penelitian menunjukan bahwa prediksi harga saham menggunakan RNN mendapatkan hasil kurang baik. Jumlah epoch, model optimasi dan ukuran learning rate untuk proses pelatihan dapat mempengaruhi hasil akurasi yang diperoleh	Kelemahan pada penelitian ini hanya menggunakan 2 parameter yaitu epoch dan learning rate.	Penelitian ini melakukan pengujian dengan menggunakan metode optimasi Stochastic Gradient Descent (SGD) dan Adaptive Moment Estimation (Adam), pengujian jumlah data dan juga pengujian learning rate. Penelitian yang akan dilakukan menggunakan metode SVR dan LSTM. Menggunakan metode Optimasi adam dan Rmsprop.
2	Peramalan Jumlah Kecelakaan Lalu Lintas Menggunakan Metode Support Vector Regression	Ni Putu Ratindia Apriyanti, I Ketut Gede Darma Putra, I Made Suwija Putra, JURNAL Ilmiah Merpati Vol 8 No 2, Agustus 2020	Hasil peramalan dapat dijadikan tinjauan untuk mengambil tindakan dalam mencegah jumlah kecelakaan yang terjadi oleh pihak pemerintah dan instansi terkait	Hasil peramalan yang telah dilakukan dengan mengimplementasikan Metode Support Vector Regression menggunakan kernel Polynomial memiliki hasil akurasi yang lebih baik, yaitu dengan nilai MAPE sebesar 7,95%, dibandingkan kernel RBF yang menghasilkan nilai MAPE sebesar 13,35%	Kekurangan dari penelitian ini tidak dijelaskan proses penentuan parameter yang digunakan pada kernel polynomial maupun kernel RBF yang sangat mempengaruhi hasil peramalan.	Pada penelitian yang akan dilakukan diperbandingkan 2 algoritma yaitu LSTM dan SVR. Untuk metode SVR menggunakan kernel RBF dan polynomial. Untuk Penentuan parameter akan menggunakan metode grid search agar mendapatkan nilai parameter yang optimal.

Tabel 2.1 Matriks literatur review dan posisi penelitian
 Prediksi Realisasi Investasi menggunakan metode Support Vector Regression dan Long Short Term Memory (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	Prediksi Laju Pertumbuhan Penduduk menggunakan metode Support Vector Regression (Studi Kasus : Kota Malang)	Arynda Kusuma Dewi, Repository.ub.ac, 2020	Menerapkan metode Support Vector Regression untuk memprediksi laju pertumbuhan penduduk.	Penelitian ini menghasilkan nilai MAPE dengan menggunakan kernel linear sebesar 0,0985% dan kernel Gaussian RBF sebesar 0,38192%	Kekurangan pada penelitian ini penentuan parameter pada kernel berdasarkan penelitian sebelumnya.	Untuk penelitian yang akan dilakukan penentuan parameter kernel menggunakan metode gridsearch.
4	Comparative Analysis of Wind Speed Forecasting Using LSTM and SVM	Satyam Gangwar, Vikram Bali, Ajay Kumar, EAI Endorsed Transactions on Scalable Information Systems Vol 7, No. 25, July 2019	Melakukan analisis komparatif dari kedua model yaitu (LSTM & SVM) dan memilih yang paling sesuai dan model yang efisien di bidang prediksi kecepatan angin.	Hasil yang diperoleh menyimpulkan bahwa LSTM lebih signifikan dalam teknik peramalan	Penelitian ini tidak menginformasikan jumlah data set yang telah dikumpulkan dan pembagian dataset tersebut untuk kebutuhan testing dan training.	Pada Penelitian yang akan dilakukan akan menggunakan pembagian data set training dan testing sebanyak 70% dan 30%
5	Analisis dan implementasi long short term memory neural network untuk prediksi harga bitcoin	Muhammad Wildan Putra Aldi, Jondri, Annisa Aditsania, e-Proceeding of Engineering Vol 5 No 2, Agustus 2018	Mencari kombinasi parameter terbaik, dan menganalisis performansi prediksi pada model jaringan syaraf tiruan dengan arsitektur LSTM	Pada pengujian tersebut didapatkan hasil yang terbaik yaitu dengan komposisi data latih 70% dan data uji 30%, parameter 1 pola time series, jumlah 25 neuron hidden, dan max epoch adalah 100 dengan akurasi rata-rata pada data latih 95.36% dan data testing 93.5%	Penelitian sudah baik dengan sudah dilakukan preprocessing dan adanya optimasi pada training	Penelitian yang akan dilakukan menggunakan kombinasi pembagian dataset 70% training dan 30% testing serta untuk metode LSTM parameter yang digunakan hidden layer, neuron, epoch dan optimasi adam serta Rmsprop

Tabel 2.1 Matriks literatur review dan posisi penelitian
 Prediksi Realisasi Investasi menggunakan metode Support Vector Regression dan Long Short Term Memory (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
6	Metode Long Short Term Memory dan Generalized Autoregressive Conditional Heteroscedasticity untuk Pemodelan Data Saham	Jannaton Khustia Lubisa, Iqbal Kharisudin, PRISMA, Prosiding Seminar Nasional Matematika, 2021	Mengetahui metode LSTM dan metode GARCH dalam meramalkan harga saham, dan mengetahui hasil peramalan saham dengan metode LSTM dan GARCH untuk tiga periode berikutnya berdasarkan pada model terbaik	Model LSTM terbaik dipilih dengan menentukan optimizer dan hidden neural serta melihat nilai RSME terkecil. Untuk model dengan metode GARCH dipilih dengan menguji efek heteroskedastik dan nilai AIC terendah. Hasil peramalan dari metode LSTM dan GARCH dibandingkan untuk mengetahui metode mana yang lebih efektif dan akurat	Kekurangan dari penelitian ini adalah yang pertama tidak diinformasikan hasil optimasi yang paling akurat dari 3 optimasi yang ada, kedua tidak diinformasikan metode yang paling akurat antara LSTM dan GARCH.	Penelitian yang akan dilakukan pada metode LSTM menggunakan optimasi adam dan Rmsprop dan menggunakan satu layer.
7	Multi Time Steps Prediction Dengan Recurrent Neural Network Long Short Term Memory	Ahmad Ashril Rizal dan Siti Soraya, Jurnal Matrik Vo. 18 No.1, November 2018	Tujuan dari penelitian ini mengkaji prediksi kunjungan wisatawan dengan pendekatan Recurrent Neural Network Long Short Term Memory	Model training dengan 3 model yaitu LSTM regression, LSTM sliding window dan LSTM time steps dan masing-masing hasil yang beragam dengan hasil yang paling baik dengan menggunakan model regresi.	Model training dengan LSTM regression, LSTM dengan sliding window dan LSTM dengan time steps belum mendapatkan hasil yang terbaik	Penelitian yang akan dilakukan menggunakan metode LSTM dengan model time step skenario bulanan, tahunan dan tiga tahunan.
8	Prediksi Kebutuhan Obat dengan Support Vector Regression (SVR dan Long Short Term Memory (LSTM)	Ekky Novriana Alam, Institut Teknologi Bandung, September 2019	Menghasilkan prediksi otomatis jumlah kebutuhan obat dengan LSTM atau SVR untuk dataset kecil selama 1 tahun dengan jumlah jenis obat yang banyak.	Algoritma LSTM memiliki hasil prediksi yang lebih baik pada prediksi mingguan. Secara umum algoritma memiliki hasil yang lebih baik dibandingkan dengan algoritma SVR.	Untuk meningkatkan akurasi prediksi dapat menggunakan data dan feature yang lebih banyak.	Penelitian yang akan dilakukan menggunakan algoritma SVR dengan grid search dan LSTM dengan optimasi adam dan Rmsprop. Prediksi dibagi dalam bulanan, tahunan dan tiga tahunan.

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Jenis penelitian ini merupakan penelitian tindakan dimana penelitian ini menerapkan dua tindakan yaitu melakukan prediksi realisasi investasi menggunakan metode SVR dan tindakan kedua dengan menggunakan metode LSTM sehingga akan didapat akurasi dari dua Tindakan tersebut. Sifat penelitian yang akan dilakukan adalah deskriptif karena penelitian ini dilakukan untuk menjelaskan karakteristik objek pada kondisi tertentu. Pendekatan kuantitatif digunakan oleh peneliti dimana penelitian akan melakukan penelitian sesuai alur yang telah peneliti buat dan hasil yang dihasilkan berupa angka akurasi prediksi nilai realisasi investasi.

3.2. Metode Pengumpulan Data

Metode pengumpulan data dilakukan dengan observasi langsung terhadap data realisasi investasi di DPMPSTP Kota Magelang dari bulan Juli tahun 2007 sampai tahun 2020 sejumlah 162. Kemudian dilakukan pemilahan sesuai data yang dibutuhkan, yaitu hanya data realisasi investasi PMDN dari tahun 2007 sampai tahun 2020.

3.3. Metode Analisis Data

Analisis data pada penelitian ini dilakukan dengan teknik prediksi menggunakan algoritma Support Vector Regression (SVR) dan algoritma Long Short Term Memory (LSTM). Pada algoritma SVR akan membandingkan dua kernel yaitu kernel RBF dan kernel polynomial, sedangkan pada algoritma LSTM akan membandingkan dua optimasi yaitu optimasi Adam dan Rmsprop. Sehingga akan diperoleh perbandingan hasil prediksi dari masing-masing algoritma tersebut untuk mendapatkan algoritma yang terbaik dalam melakukan prediksi.

3.4. Alur Penelitian

Penelitian ini dilakukan dengan tahapan sebagai berikut :

1. Pengumpulan data dilakukan dengan cara observasi langsung terhadap data realisasi investasi untuk mengidentifikasi data yang dibutuhkan dalam penelitian ini. Pemilahan data dan variabel yang akan diteliti yaitu realisasi investasi per bulan dari tahun 2007-2020, kemudian data tersebut akan dikonsultasikan dengan petugas pegawai penanaman modal.
2. Processing data dilakukan dengan melakukan cleaning data dan normalisasi data
3. Prediksi data realisasi investasi yang akan dibagi menjadi 3 skenario yaitu : prediksi jangka pendek dalam waktu 1 bulan untuk melakukan prediksi 1 bulan berikutnya, prediksi jangka menengah dalam kurun waktu 12 bulan untuk melakukan prediksi bulan ke 13 dan prediksi jangka panjang dalam waktu 3 tahun atau 36 bulan untuk melakukan prediksi bulan ke 37 dan seterusnya.

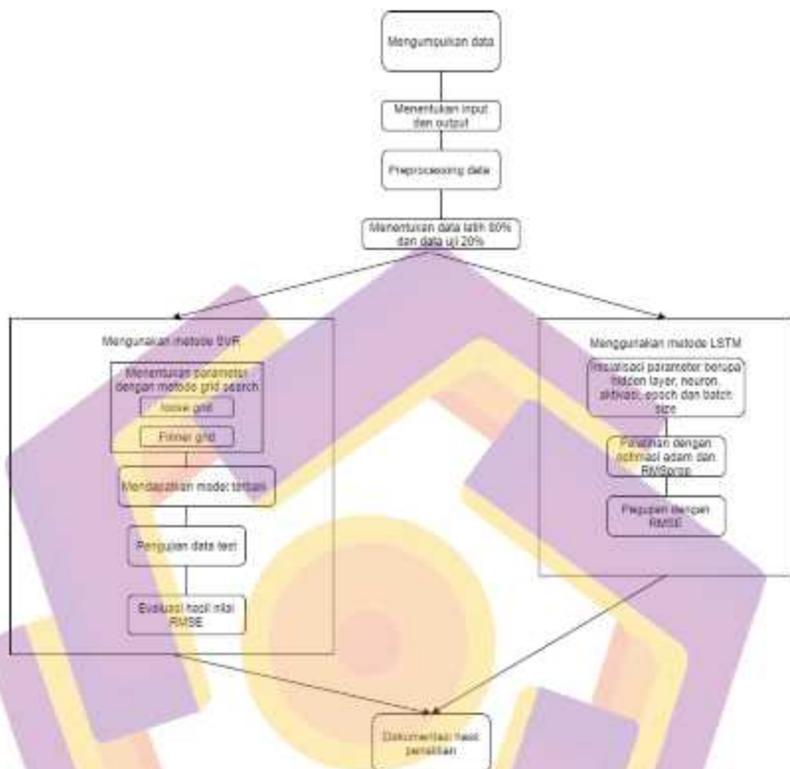
4. Pembagian data menjadi set data latih dan set data uji.
5. Proses prediksi dengan algoritma SVR dan LSTM.

Langkah-langkah dalam menggunakan SVR adalah:

- a. Menentukan parameter kemudian agar optimal menggunakan metode grid search yang dibagi dengan dua tahap yaitu loose grid dan finer grid untuk digunakan pada masing-masing kernel.
- b. Memperoleh model SVR terbaik
- c. Melakukan evaluasi terhadap hasil prediksi dengan menggunakan RMSE

Langkah-langkah dalam metode LSTM adalah :

- a. Menentukan parameter yang akan digunakan terdiri dari hidden layer, jumlah neuron, aktivasi, batch size dan epoch untuk memperoleh model LSTM terbaik
- b. Melakukan optimasi dengan metode Adam dan Rmsprop
- c. Melakukan evaluasi terhadap hasil prediksi dengan menggunakan RMSE
6. Membandingkan tingkat akurasi hasil pengolahan data menggunakan SVR dan LSTM.



Gambar 3.1 Alur Penelitian

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Gambaran Umum Obyek Penelitian

4.1.1 Profil Obyek Penelitian

Penyelenggaraan PPTSP di Kota Magelang diawali dengan dibentuknya Unit Pelayanan Tehnis Daerah (UPTD) Pelayanan Satu Atap (PSA) yang merupakan bagian dari Dinas Pendapatan Daerah Kota Magelang, sebagai cikal bakalnya.

Dengan terbitnya Peraturan Menteri Dalam Negeri Nomor 24 Tahun 2006 tentang Pedoman Penyelenggaraan Pelayanan Terpadu Satu Pintu maka Pemerintah Kota Magelang membentuk Dinas Pelayanan Terpadu Dan Penanaman Modal (DPTPM) berdasarkan Peraturan Daerah Kota Magelang Nomor 1 Tahun 2007 tentang Perubahan atas Peraturan Daerah Nomor 5 Tahun 2003 tentang Susunan Organisasi dan Tata Kerja Dinas Daerah, yang diluncurkan pada tanggal 28 Juni 2007.

Pemerintah Kota Magelang kembali melakukan penyesuaian nomenklatur untuk DPTPM berdasarkan pada Peraturan Pemerintah Nomor 41 Tahun 2007 tentang Organisasi Perangkat Daerah dan Peraturan Menteri Dalam Negeri Nomor 20 Tahun 2008 sehingga terbentuk Unit Pelayanan Perizinan Terpadu, dibentuk Badan Pelayanan Perizinan Terpadu (BP2T) Kota Magelang, yang dituangkan dalam Peraturan Daerah Kota Magelang Nomor 5 Tahun 2008 tentang Kedudukan dan Tugas Pokok Organisasi Lembaga Teknis Daerah, BP2T dan Satuan Polisi Pamong Praja, yang pejabat strukturalnya dilantik pada bulan Agustus 2008. Uraian

tentang tugas pokok dan fungsi BP2T diatur dalam Peraturan Walikota Magelang Nomor 22 Tahun 2008 tentang Penjabaran Tugas Pokok, Fungsi dan Rincian Tugas Jabatan Struktural di Lingkungan Lembaga Teknis Daerah, Badan Pelayanan Perizinan Terpadu dan Satuan Polisi Pamong Praja.

Pemerintah kembali melakukan penyesuaian untuk mengoptimalkan peran BP2T dengan terbitnya Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 15 Tahun 2014 tentang Pedoman Standar Pelayanan dan Peraturan Menteri Dalam Negeri Nomor 100 Tahun 2016 tentang Pedoman Nomenklatur Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu Provinsi dan Kabupaten/Kota. Hal tersebut diikuti dengan ditetapkannya Peraturan Daerah Kota Magelang Nomor 3 Tahun 2016 tentang Pembentukan dan Susunan Perangkat Daerah Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP) Kota Magelang. Dengan terbentuknya DPMPTSP tersebut diharapkan peran dan tugas utama terkait penanaman modal dapat dilaksanakan dengan lebih maksimal.

4.1.2 Moto Organisasi

“Siap Melayani Anda”

4.2 Pengumpulan Data

Data realisasi investasi penanaman modal didapatkan dari Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu Kota Magelang yaitu pada Bidang Pengendalian Pelaksanaan Penanaman Modal. Data realisasi Investasi dikelola oleh petugas khusus pada Bidang Pengendalian Pelaksanaan Penanaman Modal dimana data tersebut diambil dari data pengajuan izin usaha dan pelaporan

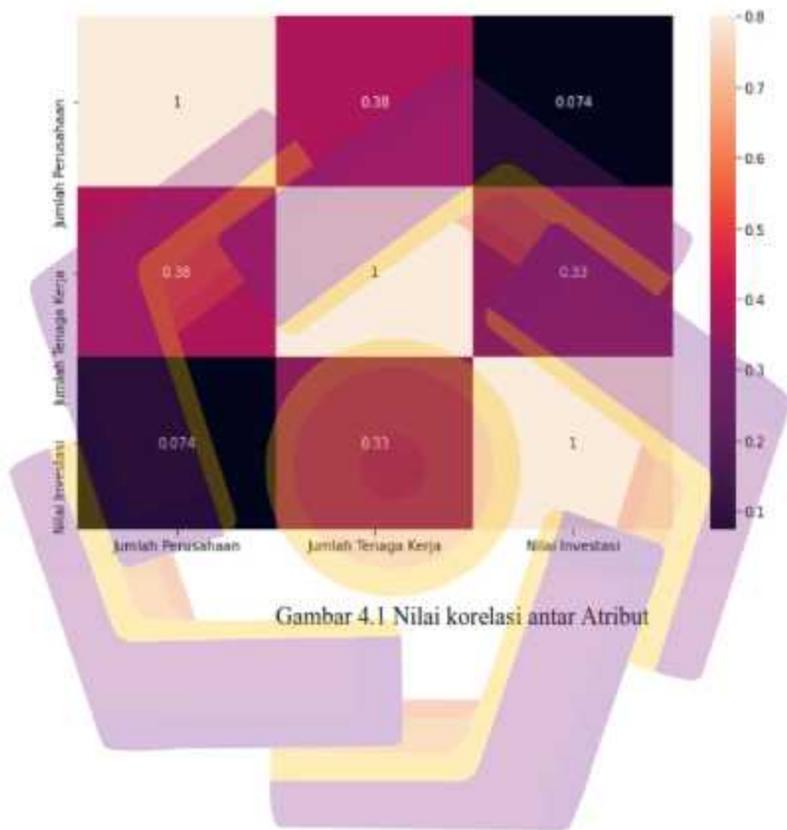
perkembangan penanaman modal yang dilakukan oleh pengusaha. Total nilai realisasi investasi akan direkap setiap bulannya dan dilaporkan sebagai perkembangan realisasi investasi. Data realisasi investasi tersebut terdiri dari 5 variabel yaitu Tanggal, jumlah perusahaan, jumlah tenaga kerja, jenis penanaman modal dan Realisasi Investasi. Kemudian dilakukan pemilahan data yang digunakan yaitu data realisasi Penanaman Modal Dalam Negeri dari bulan Juli 2007 sampai bulan Desember 2020 sehingga variabel jenis penanaman modal kita digunakan karena sudah ditentukan jenis penanaman modal yang akan digunakan. Sehingga jumlah atribut pada data realisasi investasi ada 3 atribut. Data merupakan data bulanan dengan jumlah data 162. Keterangan dan tipe data masing-masing atribut terdapat pada tabel 4.1

Tabel 4.1 Daftar Atribut Data Realisasi Investasi

No	Atribut	Keterangan	Tipe Data
1	Tanggal	Tanggal laporan	Number
2	Jumlah Perusahaan	Jumlah Perusahaan	Number
3	Jumlah Tenaga Kerja	Jumlah Tenaga Kerja	Number
4	Nilai Investasi	Jumlah Investasi	Number

Dari ketiga atribut tersebut setelah dilakukan cek relasinya didapatkan hasil bahwa atribut jumlah perusahaan memiliki relasi yang lebih dekat dengan tenaga kerja dengan nilai relasi 0.52 dibandingkan dengan nilai investasi yang memiliki nilai relasi yang lebih kecil dengan kedua atribut lainnya yaitu 0.33 dengan jumlah tenaga kerja dan 0.074 dengan jumlah perusahaan. Berdasarkan nilai relasi tersebut

maka pada penelitian ini akan menggunakan atribut tunggal nilai investasi. Hasil cek relasi antar atribut dapat dilihat pada gambar 4.1.



Dataset dengan menggunakan atribut tunggal nilai investasi dapat dilihat pada gambar 4.2

Nilai Investasi	
Tanggal	
01/08/2007	8.479363e+09
01/09/2007	2.962197e+09
01/10/2007	3.332520e+09
01/11/2007	2.653120e+09
01/12/2007	1.196730e+09
...	
01/09/2020	2.409714e+10
01/10/2020	8.708015e+10
01/11/2020	3.337955e+09
01/12/2020	1.316596e+10
01/01/2021	3.039055e+10

162 rows x 1 columns

Gambar 4.2 Dataset Awal Nilai Realisasi Investasi

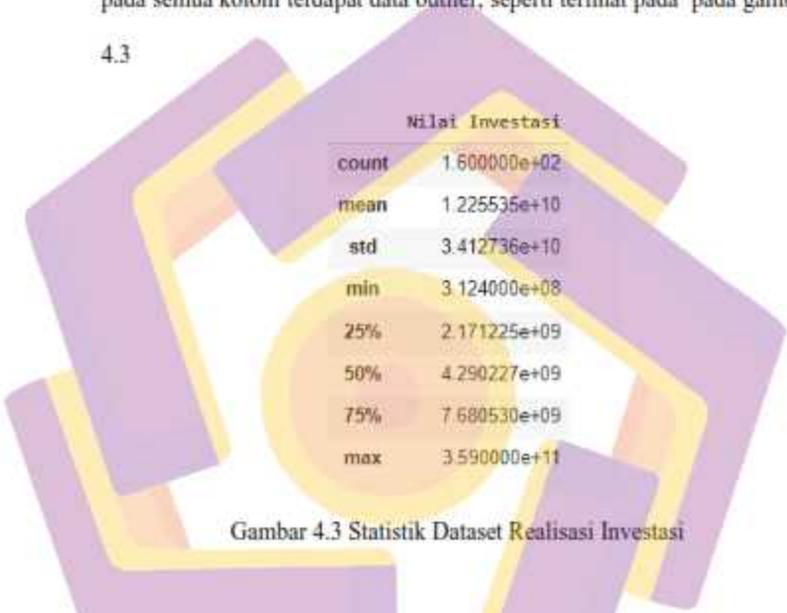
4.2.1 Preprocessing Data

Data merupakan bahan utama yang akan diolah untuk menghasilkan informasi yang berguna untuk pengambilan kebijakan. Kualitas data menjadi sangat penting untuk menghasilkan informasi yang berkualitas. Pada proses preprocessing data dilakukan dalam dua tahap yaitu mengelola data kosong dan transformasi data (Soffa, 2019). Tahap preprocessing data adalah :

1. Pembersihan Data

Mengelola dataset yang memiliki data kosong dan pada dataset yang akan digunakan ditemukan dua data kosong yaitu bulan maret 2015 dan bulan Desember 2017. Data kosong tersebut akan diisi dengan nilai rata-rata

hari sebelumnya (Soffa, 2019) namun pada penelitian ini data kosong tersebut akan dihapus sehingga tidak akan didapati nilai investasi yang bernilai null sehingga jumlah data yang digunakan menjadi 160 data. Selain data kosong pada penelitian ini juga akan menangani data outlier karena pada semua kolom terdapat data outlier, seperti terlihat pada pada gambar 4.3



Nilai Investasi	
count	1.600000e+02
mean	1.225535e+10
std	3.412736e+10
min	3.124000e+08
25%	2.171225e+09
50%	4.290227e+09
75%	7.680530e+09
max	3.590000e+11

Gambar 4.3 Statistik Dataset Realisasi Investasi

Dari gambar diatas dapat dilihat bahwa nilai maksimal memiliki selisih yang sangat jauh dengan nilai pada quartil ketiga. Maka akan diidentifikasi outlier yang memiliki nilai ekstrim yang merupakan mayor outlier (Gusti, 2014) dimana nilai tersebut akan dihapus karena akan sangat mempengaruhi model prediksi. Jumlah mayor outlier yang teridentifikasi sejumlah 7 data sehingga data setelah dilakukan penghapusan berjumlah 153 data. Setelah dilakukan wawancara dengan petugas didapatkan informasi bahwa data outlier tersebut disebabkan oleh adanya kondisi atau

kejadian khusus seperti peningkatan intensitas pemantauan, kebijakan baru, dan faktor pemicu lainnya. Karena data outlier tersebut bukan merupakan akibat dari kesalahan maka sisa data outlier yang masuk dalam kategori minor outlier akan diganti dengan nilai median supaya data tersebut tidak mempengaruhi hasil prediksi (Rya, 2017). Dataset yang sudah dilakukan preprocessing dapat dilihat pada gambar 4.4

Tanggal	Nilai Investasi
01/08/2007	8.479363e+09
01/09/2007	2.962197e+09
01/10/2007	3.332520e+09
01/11/2007	2.653128e+09
01/12/2007	1.196788e+09
...	...
01/08/2020	1.399284e+09
01/09/2020	4.055050e+09
01/11/2020	3.337955e+09
01/12/2020	1.316596e+10
01/01/2021	4.055050e+09

153 rows x 1 columns

Gambar 4.4 Dataset Realisasi Investasi

2. Transformasi Data

Transformasi data dilakukan untuk membuat nilai ada pada range tertentu sehingga tidak akan dijumpai data dengan dengan nilai selisih yang terlalu jauh (Abdul, 2019). Pada banyak penelitian transformasi data

tersebut banyak menggunakan metode min-max normalization (Jannatun, 2021) dan pada penelitian ini juga akan menggunakan normalisasi minmax dengan range $(-1,1)$ (Ahmad, 2018), untuk mengantisipasi adanya perbedaan nilai yang terlalu jauh.

4.2.2. Split Data Training dan Testing

Dataset dibagi menjadi dua bagian yaitu data training dan data testing. Data training digunakan untuk melatih model yang dibuat, sedangkan data testing digunakan untuk menguji performance dari model tersebut. Prosentase pembagian data latih dan data uji adalah 80% untuk data latih dan 20% untuk data uji (Muhammad, 2018). Karena pada penelitian ini data yang digunakan merupakan data time series maka urutan data menjadi sangat penting sehingga pembagian data dilakukan sesuai urutan dari keseluruhan data set yaitu dari bulan Juli 2007 sampai Desember 2020 dengan komposisi sebesar 80% untuk data training dan 20% untuk data testing.

4.3 Implementasi Pembentukan Model

Pada tahap ini akan dilakukan implementasi pembentukan model prediksi dengan dua algoritma pembelajaran yaitu Support Vector Regression (SVR) dan Long Short Term Memory (LSTM) sehingga dihasilkan model prediksi dari masing-masing pembelajaran. Tahap pembentukan model prediksi dibagi dalam 3 skenario yaitu :

1. Prediksi Jangka Pendek bulanan

Skenario ini dilakukan dengan data input x_{i-1} dan data output x_i karena prediksi bulanan akan melakukan prediksi bulan berikutnya.. Data training sejumlah 121 data dan data testing sejumlah 31 data.

2. Prediksi Jangka Menengah tahunan

Skenario ini dilakukan dengan data input x_{i-12}, \dots, x_{i-1} dan data output x_i karena prediksi tahunan akan melakukan prediksi 12 bulan kedepan.. Data training sejumlah 112 data dan data testing sejumlah 29 data.

3. Prediksi Jangka Panjang 3 tahunan

Skenario ini dilakukan dengan data input x_{i-36}, \dots, x_{i-1} dan data output x_i karena prediksi tiga tahunan akan melakukan prediksi 36 bulan kedepan.. Data training sejumlah 93 data dan data testing sejumlah 24 data.

4.3.1 Tahap Implementasi pembentukan model Support Vector Regression (SVR)

Pada tahap pembentukan model ini diawali dengan menentukan parameter yang akan digunakan dalam model, untuk algoritma SVR sendiri akan membandingkan dua kernel yaitu kernel RBF dan kernel polynomial. Untuk mendapatkan model yang baik dipengaruhi oleh parameter yang akan digunakan. Oleh karena itu penentuan parameter akan dilakukan menggunakan parameter grid search dengan 2 tahap yaitu tahap loose grid dan finer grid. Pada tahap loose grid akan menggunakan nilai pangkat bulat, sedangkan pada tahap finer grid menggunakan nilai disekitaran hasil terbaik yang didapatkan pada tahap loose grid

(Drajat, 2020). Pada penelitian ini parameter yang akan ditentukan adalah parameter C, gamma, epsilon dan degree dan coef atau berjumlah 5 parameter karena membandingkan 2 kernel yaitu RBF dan polynomial dimana masing-masing parameter memiliki peran dalam menghasilkan model yang terbaik. Untuk parameter yang akan ditentukan menggunakan metode search grid untuk parameter C dan gamma sedangkan untuk parameter epsilon, degree dan coef akan langsung ditentukan nilainya yaitu $\epsilon = 0.1$, $\text{degree} = 1$ dan $\text{coef0} = 0.01$.

Untuk tahap loose grid akan menggunakan range nilai yang dapat dilihat pada tabel berikut :

Tabel 4.2 Rentang nilai parameter pada tahap loose grid

Parameter	Value Range
C	$2^{-4}, 2^{-2}, \dots, 2^0, 2^4$
γ	$2^{-4}, 2^{-2}, \dots, 2^2, 2^4$

Proses pemilihan parameter menggunakan library keras dimana didalamnya dilakukan cross validation terhadap seluruh parameter yang akan di uji, karena terdapat 2 parameter yang akan diseleksi maka akan dilakukan kombinasi pencarian sebanyak jumlah parameter yaitu : 7 parameter C, 7 parameter gamma. Sehingga kombinasi parameter ada 49 kombinasi yang pada penelitian ini akan dilakukan cross validasi (Aulia, 2019) dan pada penelitian ini akan dilakukan cross validasi sebanyak 2 kombinasi data training dan testing. Parameter terbaik yang dihasilkan pada setiap skenario akan berbeda karena jumlah data dari masing-masing skenario

berbeda. Untuk rentang nilai pada tahap finer grid pada skenario satu, dua dan tiga dapat dilihat pada tabel berikut :

Tabel 4.3 Rentang nilai parameter tahap finer grid skenario bulanan

Parameter	Value Range
C	$2^{-5}, 2^{-4.75}, \dots, 2^{-1.5}, 2^{-1.25}$
γ	$2^1, 2^{1.25}, \dots, 2^{4.5}, 2^{4.75}$

Tabel 4.4 Rentang nilai parameter tahap finer grid untuk skenario tahunan

Parameter	Value Range
C	$2^{-5}, 2^{-2.75}, \dots, 2^{-0.5}, 2^{-0.75}$
γ	$2^{-1}, 2^{-0.75}, \dots, 2^{0.5}, 2^{0.75}$

Tabel 4.5 Rentang nilai parameter tahap finer grid untuk skenario tiga tahunan

Parameter	Value Range
C	$2^{-3}, 2^{-2.75}, \dots, 2^{-0.5}, 2^{-0.75}$
γ	$2^1, 2^{1.25}, \dots, 2^{2.5}, 2^{2.75}$

Pada tahap finer grid juga menggunakan library skit learn dengan proses yang sama seperti pada tahap loose grid. Jumlah kombinasi parameter yaitu 8 parameter C, 8 parameter gamma sehingga jumlah kombinasi 64 kombinasi. Hasil pada tahap finer grid setelah dilakukan cross validation dengan 2 kombinasi data training dan data testing pada masing-masing skenario yaitu :

Tabel 4.6 Nilai parameter terbaik pada seluruh skenario

Parameter	Bulanan	Tahunan	Tiga tahunan
C	$2^{-3.25}$	$2^{-2.25}$	$2^{-2.75}$
γ	$2^{4.75}$	$2^{-0.25}$	$2^{1.5}$

Nilai parameter terbaik yang dihasilkan dengan menggunakan metode gridsearch memiliki nilai yang beragam pada masing-masing skenario, bergantung dari jumlah data maupun jumlah input yang digunakan. Untuk parameter yang digunakan pada algoritma SVR dengan kernel RBF maupun polynomial pada skenario bulanan, tahunan dan skenario tiga tahunan dapat dilihat pada tabel berikut ini :

Tabel 4.7 Parameter model prediksi algoritma SVR skenario bulanan

Kernel	Parameter
RBF	$C = 2^{-3.25}$, $\gamma = 2^{4.75}$, $\epsilon = 0.1$
Polynomial	$C = 2^{-3.25}$, $\gamma = 2^{4.75}$, $\text{degree} = 1$, $\text{coef0} = 0.01$, $\epsilon = 0.1$

Tabel 4.8 Parameter model prediksi algoritma SVR skenario tahunan

Kernel	Parameter
RBF	$C = 2^{-2.25}$, $\gamma = 2^{-0.25}$, $\epsilon = 0.1$
Polynomial	$C = 2^{-2.25}$, $\gamma = 2^{-0.25}$, $\text{degree} = 1$, $\text{coef0} = 0.01$, $\epsilon = 0.1$

Tabel 4.9 Parameter model prediksi algoritma SVR skenario tiga tahunan

Kernel	Parameter
RBF	$C = 2^{-2.75}$, $\gamma = 2^{1.5}$, $\epsilon = 0.1$
Polynomial	$C = 2^{-2.75}$, $\gamma = 2^{1.5}$, $\text{degree} = 1$, $\text{coef0} = 0.01$, $\epsilon = 0.1$

Setelah mendapatkan parameter yang digunakan kita akan mendapatkan model yang akan digunakan dalam pengujian.

4.3.2 Tahap Implementasi pembentukan model Long Short Term Memory (LSTM)

Pada tahap algoritma LSTM model menggunakan library keras selain itu model akan dioptimasi menggunakan Adam dan Rmsprop yang hasilnya akan dibandingkan nilai RMSE yang didapat. Parameter yang akan digunakan untuk membuat model ditentukan langsung dengan melihat hasil penelitian sebelumnya yang menggunakan parameter 1 hidden layer, 25 neuron dan epoch 100 (Wildan, 2018), penelitian lain juga menggunakan parameter 1 hidden layer, 50 neuron (Soffa, 2017). Pada penelitian ini juga akan menggunakan parameter tersebut seperti dapat dilihat pada tabel berikut :

Tabel 4.10 Parameter model prediksi algoritma LSTM

Optimasi	Parameter
Adam	Hidden layer = 1, aktivasi = relu, Jumlah neuron = 5, epoch = 100 dan batch size = 1
Rmsprop	Hidden layer = 1, aktivasi = relu, Jumlah neuron = 5, epoch = 100 dan batch size = 1

Setelah didapatkan model prediksi dari masing-masing algoritma maka model tersebut akan disimpan untuk dipergunakan pada proses prediksi selanjutnya. Model prediksi akan digunakan pada masing-masing skenario dari bulanan, tahunan dan tiga tahunan dan akan dicari nilai RMSE terkecil. Hasil dari tesis ini adalah model yang terbaik yang dapat digunakan pada sistem prediksi nilai realisasi investasi.

4.4 Pengujian Model Prediksi

Pada tahap ini dilakukan pengujian untuk melihat akurasi hasil prediksi realisasi investasi dari model yang sudah dilatih dengan menggunakan data training.

4.4.1 Prediksi Jangka Pendek Bulanan

Berikut adalah data pada pengujian :

1. Data input : 1 atribut dengan jumlah data 121 data
2. Data Output : 1 atribut dengan jumlah data 121 data
3. Data Testing: 31 data

4.4.1.2 Hitungan manual untuk algoritma SVR prediksi bulanan adalah sebagai berikut :

- 1) Menetapkan parameter yang akan digunakan yang merupakan hasil grid search, serta nilai a_i dan a_i^* bernilai 0.
- 2) Menentukan input yang digunakan untuk memprediksi yaitu nilai investasi yang merupakan 1 data bulan sebelumnya untuk memprediksi bulan berikutnya.

Tabel 4.11 Dataset skenario bulanan

Tanggal	X_i	Target
01/08/2007	8,479,362,644	2,962,197,108
01/09/2007	2,962,197,108	3,332,520,482
01/10/2007	3,332,520,482	2,653,128,000
01/11/2007	2,653,128,000	1,196,788,315
.....
01/01/2021	13,165,956,000	30,398,647,011

3) Normalisasi data

Perhitungan normalisasi data menggunakan persamaan 2.25 sebagai berikut :

$$X_i' = \left(\frac{8,479,362,644 - 312400000}{15600179095 - 312400000} \right) \times 2 - 1$$

$$= 0.06843$$

Tabel 4.12 Normalisasi Data Bulanan

Tanggal	Nilai Investasi	y
01/08/2007	-0.205393	-0.742187
01/09/2007	-0.742187	-0.706156
01/10/2007	-0.706156	-0.772258
01/11/2007	-0.772258	-0.913953
.....
01/12/2017	-0.17489426	-0.62945074

4) Perhitungan training SVR

Untuk tahap training dilakukan dalam beberapa tahap dengan menggunakan data pada tabel 4.9 antara lain :

1. Menghitung jarak antar data training, pada perhitungan ini untuk kernel RBF jarak data ditentukan dari selisih data sedangkan pada kernel polynomial dengan mengalikan antar data.

Untuk kernel RBF menghitung selisih antar data

$$\begin{aligned} \text{Jarak} &= |-0.205393 - (-0.742187)|^2 \\ &= 0.288148 \end{aligned}$$

Tabel 4.13 Jarak antar data training bulanan RBF

Tanggal	01/08/2007	01/09/2007	01/12/2017
01/08/2007	0	0.288148	0.000930
01/09/2007	0.288148	0	0.321821
01/10/2007	0.250764	0.001298	0.282239
.....
01/12/2017	0.000930	0.321821	0

Untuk kernel Polynomial menghitung perkalian antar data training

$$\begin{aligned} \text{Jarak} &= |-0.205393 \times (-0.205393)| \\ &= 0.042186 \end{aligned}$$

Tabel 4.14 Jarak antar data training bulanan Polynomial

Tanggal	01/08/2007	01/09/2007	01/12/2017
01/08/2007	0.042186	0.152440	0.035922
01/09/2007	0.152440	0.550842	0.129804
01/10/2007	0.145039	0.524100	0.123503
.....
01/12/2017	0.035922	0.129804	0.030588

2. Menghitung matrik hessian training

Perhitungan matrik hessian menggunakan jarak data yang sudah dihitung sebelumnya ditambahkan dengan nilai lamda sesuai dengan persamaan 2.12

Untuk kernel RBF :

$$\begin{aligned} R_{12} &= \exp(-26.90 \times 0) + 0.1^2 \\ &= 1.01 \end{aligned}$$

Tabel 4.15 Matrik Hessian data training bulanan RBF

Tanggal	01/08/2007	01/09/2007	01/12/2017
01/08/2007	1.01	0.010429	0.985281
01/09/2007	0.010429	1.01	0.010173
01/10/2007	0.011174	0.975670	0.010503
.....
01/12/2017	0.985281	0.010173	1.01

Untuk kernel Polynomial

$$R_{12} = (0.042186 + 0.105)^1 + 0.1^2$$

$$= 0.157298$$

Tabel 4.16 Matrik Hessian data training bulanan Polynomial

Tanggal	01/08/2007	01/09/2007	01/12/2020
01/08/2007	0.157298	0.267552	0.151034
01/09/2007	0.267552	0.665954	0.244916
01/10/2007	0.260151	0.639212	0.238615
.....
01/12/2020	0.151034	0.244916	0.145700

3. Melakukan perhitungan sequential learning

Untuk perhitungan sequential meliputi 3 tahapan yaitu :

a. Menghitung nilai error (E_i)

Nilai yang digunakan dalam perhitungan ini adalah nilai $Var(t)$ pada tabel 4.10 dan nilai matrik hessian maupun nilai a_i dan a_i^* .

Perhitungan menggunakan persamaan 2.13 sebagai berikut :

$$E_i = -0.742187 - [(0 - 0) * 1.01]$$

$$= -0.742187$$

Tabel 4.17 Nilai error bulanan

E1
-0.742187
-0.706156
-0.772258
.....
-0.629450

b. Perubahan nilai langrange multiplier

Pada perhitungan menggunakan nilai error, gamma, epsilon, kompleksitas dan persamaan yang digunakan adalah 2.14 dan 2.15

$$\begin{aligned} \delta\alpha_i &= \min(\max[26,90(-0,742187 - 0,1), 0], 0,105 - 0) \\ &= 0,10511 \end{aligned}$$

$$\begin{aligned} \delta\alpha_i &= \min(\max[26,90(0,742187 - 0,1), -0], 0,105 - 0) \\ &= 0 \end{aligned}$$

Tabel 4.18 Perubahan Nilai langrange Multiplier Bulanan

α_1^*	α_1
0	0.10511
0	0.10511
0	0.10511
.....
0	0.10511

c. Update nilai langrange multiplier

Nilai yang digunakan pada perhitungan update langrange multiplier adalah nilai awal α_1^* dan α_1 adalah nol, serta menggunakan nilai langrange multiplier pada tabel 4.16 menggunakan persamaan 2.16 dan 2.17.

Tabel 4.19 Update Nilai Langrange multiplier Bulanan

a_1^*	a_1
0	0.10511
0	0.10511
0	0.10511
.....
0	0.10511

Tabel 4.19 merupakan hasil perhitungan dari sequential learning pada iterasi 1 dan bila dijadikan dalam satu tabel adalah sebagai berikut :

Tabel 4.20 Sequential Learning iterasi 1 Bulanan

E_1	$\delta\alpha_1^*$	$\delta\alpha_1$	$\alpha_1^*(\text{baru})$	$\alpha_1(\text{baru})$
-0.742187	0	0.10511	0	0.10511
-0.706156	0	0.10511	0	0.10511
-0.772258	0	0.10511	0	0.10511
.....
-0.629450	0	0.10511	0	0.10511

Perulangan sequential learning dilakukan terus menerus sampai jumlah iterasi yang ditentukan. Pada penelitian ini ditentukan 100 iterasi atau memenuhi syarat stop condition.

Tabel 4.21 Sequential Learning iterasi 100 Bulanan

E_1	$\delta\alpha_1^*$	$\delta\alpha_1$	$\alpha_1^*(\text{baru})$	$\alpha_1(\text{baru})$
1.255617	0	0.10511	0	0.10511
.....
-0.929286	0	0.10511	0	0.10511
1.399641	0	0	0	0
0.945085	0	0.10511	0	0.10511

5) Prediksi

Perhitungan prediksi menghasilkan nilai prediksi yang diukur menggunakan nilai RMSE yang diperoleh dengan menghitung nilai jarak data testing, nilai matrik hessian dan fungsi regresi. Untuk menghitung akurasi prediksi menggunakan dataset testing pada tabel 4.22

Tabel 4.22 Data Testing Bulanan

Tanggal	Nilai Investasi	y
01/01/2018	-0.62945074	-0.41842097
01/02/2018	-0.41842097	-0.76169941
01/03/2018	-0.76169941	-0.4884124
.....
01/12/2020	0.25059096	-0.63585766

a. Menghitung jarak antar data testing

Menghitung jarak antar data testing pada masing-masing kernel polynomial dan RBF dengan menggunakan data pada tabel 4.18.

Untuk kernel RBF :

$$\begin{aligned} \text{Jarak_testing} &= |-0.62945074 - (-0.41842097)|^2 \\ &= 0.044534 \end{aligned}$$

Tabel 4.23 Jarak antar data testing bulanan RBF

Tanggal	01/01/2018	01/02/2018	01/12/2020
01/01/2018	0	0.044534	0.774473
01/02/2018	0.044534	0	0.447577
01/03/2018	0.017490	0.117840	1.024732
.....
01/12/2020	0.774473	0.447577	0

Tabel 4.24 Jarak antar data testing bulanan Polynomial

Tanggal	01/01/2018	01/02/2018	01/12/2020
01/01/2018	0.396208	0.263375	- 0.157735
01/02/2018	0.263375	0.175076	-0.104853
01/03/2018	0.479452	0.318711	-0.190875
.....
01/12/2020	- 0.157735	-0.104853	0.062796

b. Menghitung Matrik Hessian

Perhitungan nilai dari matrik hessian data testing menggunakan nilai jarak data pada masing-masing kernel pada tabel 4.23 dan 4.24.

Untuk kernel RBF :

$$R_{12} = \exp(-26.90 \times 0.044534) + 0.1^2$$

$$= 0.311695$$

Tabel 4.25 Matrik Hessian Testing Bulanan RBF

Tanggal	01/01/2018	01/02/2018	01/12/2020
01/01/2018	1.010000	0.311695	0.010000
01/02/2018	0.311695	1.010000	0.010006
01/03/2018	0.634702	0.052008	0.010000
.....
01/12/2020	0.010000	0.010006	1.010000

Untuk Kernel Polynomial

$$R_{12} = (0.396208 + 0.105)^1 + 0.1^2$$

$$= 0.511320$$

Tabel 4.26 Matrik Hessian Testing Bulanan Polynomial

Tanggal	01/01/2018	01/02/2018	01/12/2020
01/01/2018	0.511320	0.378487	-0.042623
01/02/2018	0.378487	0.290188	0.010260
01/03/2018	0.594564	0.433823	-0.075763
.....
01/12/2020	-0.042623	0.010260	0.177908

c. Mengitung fungsi regresi $f(x)$

Nilai yang digunakan untuk melakukan perhitungan regresi $f(x)$ menggunakan nilai langgrange multiplier yang baru dari tabel 4.21 dan nilai matrik hessian pada tabel 4.26 dengan menggunakan persamaan 2.18.

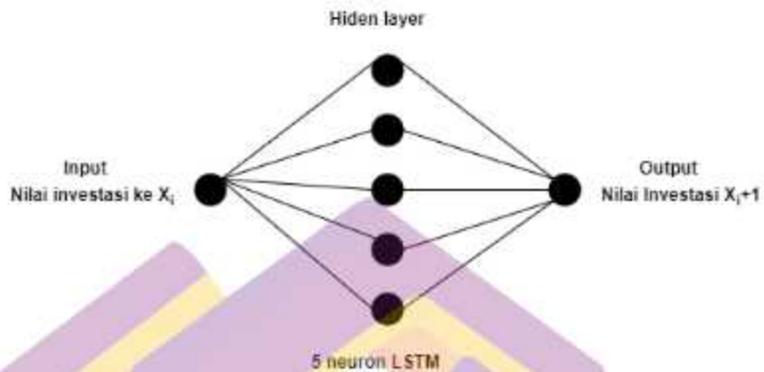
$$f(x) = [0-0.10511] \times 1.01 + [0-0] \times 0.301696..... + [0-0.10511] \times 0.01 \\ = - 0.57259$$

Tabel 4.27 Hasil Prediksi Bulanan

Aktual	Prediksi RBF	Prediksi Polynomial
-0.418421	-0.57259549	-0.71098475
-0.761699	-0.62631616	-0.65144208
.....
-0.635858	-0.73615465	-0.46267839

4.4.1.2 Hitungan manual untuk algoritma LSTM adalah sebagai berikut :

Proses pada algoritma LSTM skenario prediksi bulanan menggunakan 3 layer yaitu layer input dengan 1 data input, hidden layer dengan 5 neuron dan layer output dengan 1 data output. Arsitektur dari LSTM skenario bulanan dapat dilihat pada gambar 4.5.



Gambar 4.5 Arsitektur LSTM Prediksi Bulanan

Untuk perhitungan manual prediksi bulanan menggunakan algoritma LSTM diawali dengan membuat dataset prediksi bulanan pada tabel 4.11 dan dilakukan normalisasi data serta pembagian data training dan data testing. Berikut adalah contoh perhitungan algoritma LSTM dengan 1 data inputan pada 1 kali iterasi. Perhitungan untuk tanggal 1 januari 2021 dan untuk 1 kali iterasi.

1) Menentukan nilai awal dari :

$$h_{c-1} = 0$$

$$\text{bias forget gate} = 1$$

$$\text{bias input gate} = 0.5$$

$$\text{bias input } \bar{C}_t = 0$$

$$\text{Bias output gate} = 0.5$$

Sebelum menghitung forget gate kita menentukan nilai weight dari jumlah variabel. Untuk prediksi bulanan ini memiliki variabel 1 maka nilai weightnya adalah :

$$W = \left(-\frac{1}{\sqrt{1}}, \frac{1}{\sqrt{1}} \right)$$

$$= (-1, 1)$$

- 2) Menghitung forget gate dengan menggunakan persamaan 2.19 dan dataset yang digunakan pada tabel 4.34.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$= \sigma(1 * 0 + 1 * -0.629451) + 1$$

$$= \sigma(0.370549)$$

$$= \frac{1}{1 + e^{-0.370549}}$$

$$= 0.59159$$

- 3) Menghitung input gate dengan menggunakan dua persamaan yaitu 2.21 dan 2.22

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$= \sigma(1 * 0 + 1 * -0.629451) + 0.5$$

$$= \sigma(0.370549)$$

$$= 0.46768$$

$$C_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

$$= \tanh(1 * 0 + 1 * -0.629451 + 0)$$

$$= \tanh(-0.629451)$$

$$= 2 * \left(\frac{1}{1 + e^{-2 * -0.629451}} \right) - 1$$

$$= -0.55767$$

- 4) Menghitung update nilai cell state dengan menggunakan persamaan 2.23

$$C_t = f_t * C_{t-1} + i_t * C_t$$

$$\begin{aligned}
 &= 0.59159 * 0 + 0.46768 * -0.55767 \\
 &= -0.2608
 \end{aligned}$$

- 5) Tahap terakhir menghitung nilai output yang menggunakan persamaan 2.24 dan 2.25

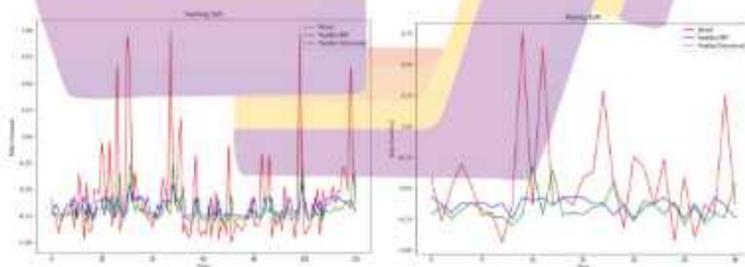
$$\begin{aligned}
 ot &= \sigma(Wo.[ht-1, xt] + bo) \\
 &= \sigma(1 * 0 + 1 * -0.629451 + 0.5) \\
 &= \sigma(-0.529451) \\
 &= 0.46768 \\
 ht &= ot * \tanh(Ct) \\
 &= 0.471468 * \tanh(-0.2608) \\
 &= 0.46768 * -0.2550 \\
 &= -0.1192
 \end{aligned}$$

Untuk nilai prediksi yang digunakan untuk dibandingkan dengan nilai aktual adalah nilai input terakhir dengan sejumlah iterasi yang telah ditentukan (Wulan, 2021). Untuk prediksi bulanan data yang digunakan sebagai data prediksi adalah data x1. Hasil pengujian Algoritma SVR dan LSTM dapat dilihat pada tabel 4.28.

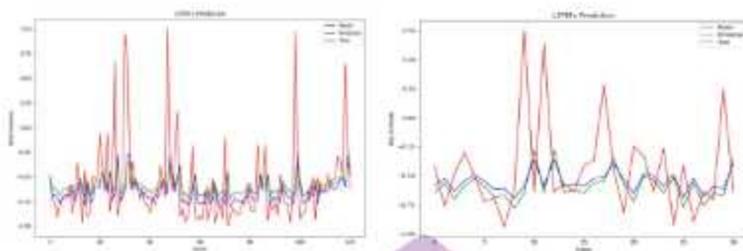
Tabel 4.28 Hasil pengujian RMSE algoritma SVR & LSTM - Bulanan

Kernel	Data Training	Data Testing
SVR -RBF	0.38077	0.44557
SVR -Polynomial	0.40534	0.49205
LSTM - Adam	0.39588	0.47118
LSTM - Rmsprop	0.38941	0.44448

Pada prediksi bulanan penggunaan kernel RBF memiliki nilai hasil prediksi yang sedikit lebih baik dibandingkan dengan kernel polynomial pada data training maupun data testing. Hasil pada data testing menggunakan kernel RBF dan polynomial memiliki nilai 0,44557 dan 0,49266 RMSE. Hasil tersebut masih lebih baik dibandingkan penelitian sebelumnya menggunakan metode yang sama yang memiliki nilai RMSE antara 198,225 dan 0,4 dengan rata-rata nilai RMSE yang didapat adalah 7,07 RMSE (Ekky,2019). Sedangkan pada pengujian menggunakan algoritma LSTM penggunaan optimasi Rmsprop memiliki nilai RMSE yang lebih baik dibandingkan dengan optimasi menggunakan Adam dengan nilai 0,44448 RMSE. Hasil tersebut masih lebih baik dibandingkan dengan penelitian sebelumnya yang memiliki nilai RMSE maksimal 57,092 dan nilai RMSE minimal 0,00152 serta nilai rata-rata MSRE 4,286 (Ekky,2019). Hasil pengujian menggunakan optimasi Rmsprop ini juga lebih baik dibandingkan dengan hasil prediksi menggunakan algoritma SVR. Gambaran hasil prediksi dibandingkan nilai aktual dapat dilihat pada gambar 4.6 dan 4.7 :



Gambar 4.6 Data training dan data testing-SVR bulanan



Gambar 4.7 Data training dan data testing- LSTM bulanan

Pada gambar tersebut dapat dilihat bahwa hasil prediksi pada data training pada algoritma SVR maupun LSTM memiliki hasil yang lebih baik dibandingkan pada data testing, walaupun hasilnya belum optimal. Pada data testing dengan menggunakan algoritma SVR pada penggunaan kernel RBF memiliki hasil prediksi yang sedikit lebih baik dari kernel Polynomial. Hal tersebut dipengaruhi oleh jumlah data yang diuji karena pada data training memiliki jumlah data yang lebih banyak dibandingkan dengan jumlah data tersting. Selain jumlah data nilai parameter yaitu *degree* pada kernel polynomial dan nilai *C* serta *gamma* juga mempengaruhi karena parameter tersebut mempengaruhi nilai langangrange multiper yang digunakan untuk pengujian pada data testing . Nilai epsilon yang besar juga mengakibatkan proses pembelajaran yang terlalu cepat sehingga hasil yang didapat menjadi tidak maksimal (Noval, 2019). Hasil prediksi pada algoritma LSTM menggunakan optimasi Rmsprop lebih baik dibandingkan dengan menggunakan optimasi Adam, walaupun keduanya belum optimal dalam melakukan prediksi. Pada prediksi bulanan ini dipengaruhi oleh jumlah inputan yang hanya memiliki satu data input sehingga data pembelajaran sangat terbatas.

4.4.2 Prediksi Jangka Menengah Tahunan

Berikut adalah data pada pengujian :

1. Data input : 12 atribut dengan jumlah data 112 data
2. Data Output : 1 atribut dengan jumlah data 112 data
3. Data Testing: 29 data

4.4.2.1 Hitungan manual untuk algoritma SVR adalah sebagai berikut :

Untuk tahapan prediksi pada skenario tahunan ini sama seperti skenario bulanan hanya saja jumlah matrik berbeda karena pada prediksi tahunan ini memiliki matrik data training sejumlah 112, dengan jumlah data input 12 data. Untuk data yang digunakan pada prediksi tahunan adalah sebagai berikut :

Tabel 4.29 Dataset skenario tahunan

Tanggal	X1	X2	Target
01/08/2008	8.479363e+09	2.962197e+09	6.929611e+09
01/09/2008	2.962.197,108	3.332.520,482	4.271134e+09
01/10/2008	3.332.520,482	2.653.128,000	6.460000e+08
.....
01/11/2019	13.165.956,000	30.398.647,011	5.885491e+09

Selanjutnya untuk penskalaan data agar berada pada range tertentu dapat dilihat pada tabel 4.30

Tabel 4.30 Normalisasi Data Tahunan

Tanggal	X1	X2	Target
01/08/2008	-0.205393	-0.742187	-0.356176
01/09/2008	-0.742187	-0.706156	-0.614834
01/10/2008	-0.706156	-0.772258	-0.967542
.....
01/11/2019	-0.457764	-0.819021	-0.635858

Setelah dilakukan normalisasi dan pembagian data training maupun data testing maka akan dilakukan perhitungan jarak antar data dan matrik hessian berdasarkan kernel polynomial dan kernel RBF. Karena memiliki 12 data input maka perhitungan jarak antar data akan menghitung seluruh data input. Nilai dari jarak antar data dan matrik hessian dapat dilihat pada tabel 4.31 dan 4.32.

Perhitungan jarak antar data pada kernel RBF :

$$\begin{aligned} \text{Jarak antar data} &= [-0.205393 - (-0.742187)]^2 + [-0.742187 - (-0.706156)]^2 + \dots + [- \\ &\quad 0.693238 - (-0.356176)]^2 \\ &= 0.579093 \end{aligned}$$

Tabel 4.31 Jarak antar data training tahunan RBF

Tanggal	01/08/2007	01/09/2007	01/12/2016
01/08/2007	0	0.579093	3.106085
01/09/2007	0.579093	0	2.526993
01/10/2007	0.656906	0.077814	2.449179
.....
01/12/2016	3.106085	2.526993	0

Perhitungan jarak antar data pada kernel Polynomial :

$$\begin{aligned} \text{Jarak} &= |-0.205393 \times -0.205393| + |-0.706156 \times -0.706156| + \dots + |-0.356176 \\ &\quad \times -0.356176| \\ &= 0.042186 \end{aligned}$$

Tabel 4.32 Jarak antar data training tahunan Polynomial

Tanggal	01/08/2007	01/09/2007	01/12/2016
01/08/2007	0	0.247209	-3.197627
01/09/2007	0.247209	0	-2.950418
01/10/2007	0.372526	-0.125318	2.825100
.....
01/12/2016	-3.197627	-2.950418	0

Untuk menghitung matriks hessian menggunakan nilai jarak data pada masing-masing kernel. Untuk hasil perhitungan dapat dilihat pada tabel 4.33 dan 4.34

Untuk kernel RBF

$$R_{12} = \exp(-0.8408 \times 0.579093) + 0.1^2$$

$$= 0.123265$$

Tabel 4.33 Matrik Hessian RBF

Tanggal	01/08/2007	01/09/2007	01/12/2016
01/08/2007	1.010000	0.010000	0.010000
01/09/2007	0.010000	1.010000	0.010000
01/10/2007	0.010000	0.133292	0.010000
.....
01/12/2016	0.010000	0.010000	1.010000

Untuk kernel Polynomial

$$R_{12} = (0.247209 + 0.2102)^1 + 0.1^2$$

$$= 0.216984$$

Tabel 4.34 Matrik hessian Polynomial

Tanggal	01/08/2007	01/09/2007	01/12/2016
01/08/2007	0.220224	0.467433	-2.977403
01/09/2007	0.467433	0.220224	-2.730194
01/10/2007	0.592750	0.094906	3.045324
.....
01/12/2016	-2.977403	-2.730194	0.220224

Tahapan penghitungan sequential learning :

1. Melakukan penghitungan nilai error (E_1)

Untuk menghitung nilai error menggunakan nilai pada kolom target pada tabel 4.30, dengan nilai awal α_i dan α_i^* adalah nol.

$$E_i = -0.356176 - [(0 - 0) * 1.01] + (0 - 0) * 0.01 + \dots + (0 - 0) * 0.01$$

$$= -0.356176$$

Tabel 4.35 Nilai Error Tahunan

E1
-0.356176
-0.614834
-0.967542
.....
-0.635858

2. Melakukan perubahan nilai langgrange multiplier

Pada perhitungan menggunakan nilai error pada tabel 4.35 dan gamma, epsilon, kompleksitas pada tabel 4.3. Persamaan yang digunakan adalah 2.14 dan 2.15

$$\delta a1^* = \min (\max [0.8408(-0.356176 - 0.1), 0], 0.2102 - 0) \\ = 0$$

$$\delta a1 = \min (\max [0.8408(0.356176 - 0.1), -0], 0.2102 - 0) \\ = 0.2102$$

Tabel 4.36 Perubahan Nilai Langgrange Multiplier Tahunan

a1*	a1
0	0.2102
0	0.2102
0	0.2102
.....
0	0.2102

3. Melakukan update

Nilai yang digunakan pada perhitungan update langrange multiplier adalah nilai awal a1* dan a1 adalah nol, serta menggunakan nilai langrange multiplier pada tabel 4.36 menggunakan persamaan 2.16 dan 2.17.

Tabel 4.37 Update Nilai Langrange Multipler Tahunan

$a1^*$	$a1$
0	0.2102
0	0.2102
0	0.2102
.....
0	0.2102

Untuk hasil perhitungan sequential learning pada iterasi 1 bila dijadikan satu tabel adalah :

Tabel 4.38 Sequential Learning 1 Iterasi Tahunan

E_1	$\delta\alpha_i^*$	$\delta\alpha_i$	$\alpha_i^*(\text{baru})$	$\alpha_i(\text{baru})$
-0.356176	0	0.2102	0	0.2102
-0.614834	0	0.2102	0	0.2102
-0.967542	0	0.2102	0	0.2102
.....
-0.635858	0	0.2102	0	0.2102

Perulangan sequential learning dilakukan terus menerus sampai jumlah iterasi yang ditentukan. Pada penelitian ini ditentukan 100 iterasi atau memenuhi syarat stop condition.

Tabel 4.39 Sequential Learning 100 Iterasi Tahunan

E_1	$\delta\alpha_i^*$	$\delta\alpha_i$	$\alpha_i^*(\text{baru})$	$\alpha_i(\text{baru})$
0.0919638	0	0.2102	0	0.2102
-0.614834	0	0.2102	0	0.2102
0.1265065	0	0.2102	0	0.2102
.....
-0.635858	0	0.2102	0	0.2102

Nilai langgrange terbaru tersebut akan digunakan pada proses prediksi pada data testing. Untuk menghitung fungsi prediksi akan dilakukan

perhitungan jarak data dan matrik hessian. Berikut data testing yang digunakan :

Tabel 4.40 Dataset Testing Tahunan

Tanggal	X1	X2	Target
01/04/2017	-0.46388571	-0.58601816	-0.4884124
01/05/2017	-0.58601816	-0.51688102	-0.3006434
01/06/2017	-0.51688102	-0.52117227	-0.4923157
.....
01/11/2019	-0.45776428	-0.81902096	-0.6358576

Perhitungan jarak antar data pada masing-masing kernel RBF dan polynomial menggunakan nilai pada tabel 4.30

Untuk Kernel RBF

$$\begin{aligned} \text{Jarak_testing} &= [-0.46388571 - (-0.58601816)]^2 + [-0.58601816 - \\ &(-0.51688102)]^2 + \dots + [-0.4884124 - (-0.3006434)]^2 \\ &= 2.590410 \end{aligned}$$

Tabel 4.41 Jarak antar data testing tahunan RBF

Tanggal	01/04/2017	01/05/2017	01/11/2019
01/04/2017	0	2.590410	3.189500
01/05/2017	2.590410	0	3.552276
01/06/2017	3.117232	2.610751	2.156941
.....
01/11/2019	3.189500	3.552276	0

Untuk kernel Polynomial

$$\begin{aligned} \text{Jarak testing} &= [-0.46388571 * -0.58601816] + [-0.58601816 * 0.51688102] \\ &+ \dots + [-0.4884124 * -0.3006434] \\ &= 1.873349 \end{aligned}$$

Tabel 4.42 Jarak antar data testing tahunan Polynomial

Tanggal	01/04/2017	01/05/2017	01/11/2019
01/04/2017	3.156876	1.873349	2.145269
01/05/2017	1.873349	3.180233	1.975559
01/06/2017	1.483423	1.748341	2.546712
.....
01/11/2019	2.145269	1.975559	4.323163

Setelah mendapatkan perhitungan jarak data maka akan dihitung nilai matrik hessian sesuai dengan masing-masing kernel RBF dan Polynomial

Untuk kernel RBF

$$R_{12} = \exp(-0.8408 \times 2.590410) + 0.3^{-1} \\ = 0.203237$$

Tabel 4.43 Matrik Hessian tahunan RBF

Tanggal	01/04/2017	01/05/2017	01/11/2019
01/04/2017	1.09	0.203237	0.158423
01/05/2017	0.203237	1.09	0.140433
01/06/2017	0.162710	0.201317	0.253039
.....
01/11/2019	0.158423	0.140433	1.09

Untuk kernel Polynomial

$$R_{12} = (3.156876 + 0.21022410381342863)^1 + 0.3^2 \\ = 3.457100$$

Tabel 4.44 Matrik Hessian Tahunan Polynomial

Tanggal	01/04/2017	01/05/2017	01/11/2019
01/04/2017	3.457100	2.173573	2.445493
01/05/2017	2.173573	3.480457	2.275783
01/06/2017	1.783647	2.048566	2.846936
.....
01/11/2019	2.445493	2.275783	4.623387

Selanjutnya untuk melakukan perhitungan regresi $f(x)$ menggunakan nilai langgrange multiplier yang baru dari tabel 4.39 dan nilai matrik hessian pada tabel 4.44

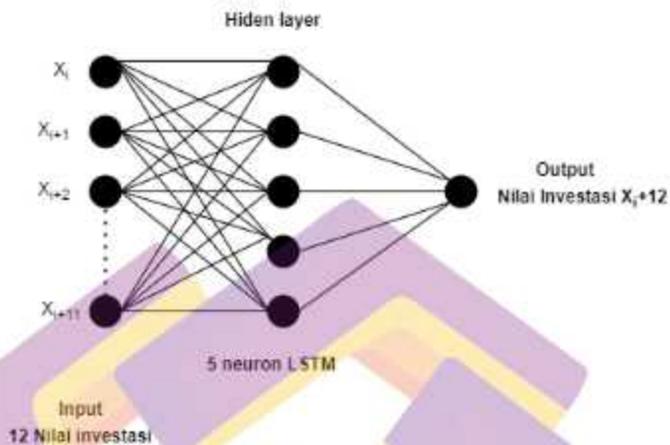
$$f(x) = [0-0.2102] \times 1.01 + [0-0] \times 0.123265..... + [0.2102] \times 0.078444 \\ = -0.57259$$

Tabel 4.45 Hasil Prediksi tahunan

Aktual	Prediksi RBF	Prediksi Polynomial
-0.4884124	-0.7205426	-6.92151
-0.30064342	-0.5292562	-6.73734
.....
-0.63585766	-0.82305168	-9.43943

4.4.2.2 Hitungan manual untuk algoritma LSTM adalah sebagai berikut :

Proses pada algoritma LSTM untuk skenario prediksi tahunan dengan 3 layer yaitu layer input dengan 12 data input, hidden layer dengan 5 neuron LSTM dan layer output dengan 1 data output. Arsitektur dari LSTM skenario tahunan dapat dilihat pada gambar 4.8



Gambar 4.8 Arsitektur LSTM Prediksi Tahunan

Untuk perhitungan algoritma LSTM diawali dengan membuat dataset times series tahunan pada tabel 4.30 dan dilakukan normalisasi data serta pembagian data training dan data testing.

Berikut adalah contoh perhitungan algoritma LSTM dengan 12 data inputan pada 1 kali iterasi untuk data tanggal 1 April 2018.

- 6) Menentukan nilai awal dari :

$$h_{t-1} = 0$$

$$\text{bias forget gate} = 1$$

$$\text{bias input gate} = 0.5$$

$$\text{bias input } \bar{C}_t = 0$$

$$\text{Bias output gate} = 0.5$$

Sebelum menghitung forget gate kita menentukan nilai weight dari jumlah variabel. Untuk prediksi tahunan ini memiliki variabel 12 maka nilai weightnya adalah :

$$W = \left(-\frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}} \right)$$

$$= (-0.288, 0.288)$$

- 7) Menghitung forget gate dengan menggunakan persamaan 2.19 dan dataset yang digunakan pada tabel 4.30.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$= \sigma(0.288 * 0 + 0.288 * -0.463886) + 1$$

$$= \sigma(0.8864)$$

$$= \frac{1}{1 + e^{-0.8864}}$$

$$= 0.70814$$

- 8) Menghitung input gate dengan menggunakan dua persamaan yaitu 2.21 dan 2.22

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$= \sigma(0.288 * 0 + 0.288 * -0.463886) + 0.5$$

$$= \sigma(0.3864)$$

$$= 0.5954$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

$$= \tanh(0.288 * 0 + 0.288 * -0.463886) + 0$$

$$= \tanh(-0.13359)$$

$$= 2 * \left(\frac{1}{1 + e^{-2 * (-0.13359)}} \right) - 1$$

$$= -0.1328$$

9) Menghitung update nilai cell state dengan menggunakan persamaan 2.23

$$\begin{aligned} C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\ &= 0.70814 * 0 + 0.5954 * -0.1328 \\ &= -0.07906 \end{aligned}$$

10) Tahap terakhir menghitung nilai output yang menggunakan persamaan 2.24 dan 2.25

$$\begin{aligned} o_t &= \sigma(W_o.[h_{t-1}, x_t] + b_o) \\ &= \sigma(0.288 * 0 + 0.288 * -0.205393) + 0.5 \\ &= \sigma(0.3864) \\ &= 0.5954 \\ h_t &= o_t * \tanh(C_t) \\ &= 0.5954 * \tanh(-0.07906) \\ &= 0.5954 * -0.07889 \\ &= -0.0469 \end{aligned}$$

Perhitungan dilakukan untuk seluruh data input dari x1 sampai x12, untuk nilai bias dan weight menggunakan nilai yang sama namun untuk nilai h_t dan C_t menggunakan nilai hasil perhitungan sebelumnya. Untuk nilai prediksi yang digunakan untuk dibandingkan dengan nilai aktual adalah nilai input terakhir dan untuk prediksi tahunan data yang digunakan sebagai data prediksi adalah data ke12 atau x12.

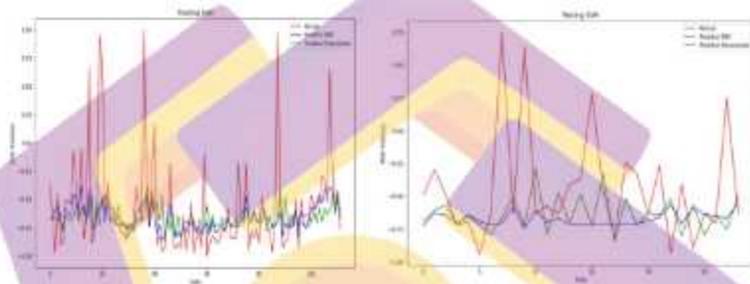
Hasil pengujian Algoritma SVR & LSTM dapat dilihat pada tabel 4.46.

Tabel 4.46 Hasil pengujian RMSE algoritma SVR & LSTM - Tahunan

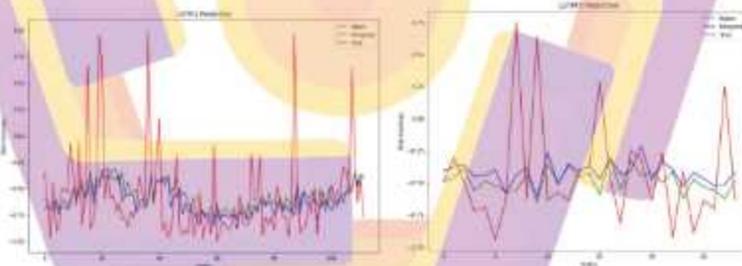
Kernel	Data Training	Data Testing
SVR - RBF	0.36142	0.50202
SVR - Polynomial	0.40680	0.50176
LSTM - Adam	0.38977	0.45134
LSTM - Rmsprop	0.39802	0.44588

Pada hasil pengujian prediksi tahunan juga memiliki hasil yang sama seperti pada prediksi bulanan yaitu hasil prediksi pada data training lebih baik dibandingkan data testing. Hasil pada data testing menggunakan algoritma SVR dengan kernel RBF dan polynomial memiliki nilai yang hampir sama yaitu 0,50202 dan 0,50176 RMSE. Hasil tersebut masih lebih baik dibandingkan nilai rata-rata RMSE pada penelitian sebelumnya dengan nilai 6,510 RMSE, namun pada nilai minimal RMSE sebesar 0.039 RMSE yang mempunyai nilai yang lebih baik daripada penelitian ini (Ekky,2019). Hal tersebut karena ada perbedaan jumlah input dan output, pada penelitian sebelumnya merupakan prediksi mingguan dengan 7 data input dan 7 data output, sedangkan pada penelitian ini merupakan prediksi tahunan dengan 12 data input dan 1 data output. Demikian juga pada pengujian menggunakan algoritma LSTM dengan optimasi Rmsprop pada prediksi tahunan juga memiliki hasil yang lebih baik dibandingkan optimasi Adam dengan nilai 0,44588 RMSE. Hasil tersebut lebih buruk bila dibandingkan dengan penelitian sebelumnya pada nilai RMSE minimal dengan nilai 0,042 RMSE namun lebih baik bila dibandingkan dengan nilai RMSE maksimal yaitu 52,724 maupun nilai RMSE rata-rata yaitu 3,618(Ekky, 2019). Hasil prediksi tahunan dengan algoritma LSTM dan optimasi

Rmsprop ini hampir sama bila dibandingkan dengan hasil prediksi bulanan, dan bila dibandingkan dengan hasil prediksi menggunakan algoritma SVR pada prediksi tahunan hasil prediksi menggunakan algoritma LSTM optimasi Rmsprop maupun Adam hasilnya tetap lebih baik. Gambaran hasil prediksi dibandingkan dengan nilai aktual dapat dilihat pada gambar 4.9 dan 4.10.



Gambar 4.9 Data Training dan data testing- SVR Tahunan



Gambar 4.10 Data Training dan data testing- LSTM Tahunan

Pada gambar tersebut dapat dilihat hasil prediksi pada data training dan data testing belum mendapatkan hasil yang optimal. Hal tersebut dipengaruhi oleh nilai gamma yang besar dibandingkan dengan nilai C, selain itu juga dipengaruhi nilai langgrange multipler yang memiliki nilai mendekati nilai C dan lebih besar dari nilai epsilon sehingga banyak data yang diluar support vektor. Selain itu nilai

gamma yang kecil menjadi kurang menangkap pola yang penting (Olivia, 2018). Pada algoritma LSTM dengan menggunakan optimasi Rmsprop mempunyai nilai hasil pengujian yang lebih baik, namun hasil tersebut juga belum optimal. Pada prediksi tahunan ini hasil prediksi dipengaruhi oleh sistem kerja optimasi dan input data yang sejumlah 12 data sehingga data pembelajaran lebih banyak.

4.4.3. Prediksi Jangka Panjang Tiga Tahunan

Berikut adalah data pada pengujian :

1. Data input : 36 atribut dengan jumlah data 100 data
2. Data output : 1 atribut dengan jumlah data 100 data
3. Data testing : 26 data

4.4.3.1 Hitungan manual untuk algoritma SVR adalah sebagai berikut :

Untuk tahapan prediksi sama seperti skenario bulanan maupun tahunan, hanya jumlah matrik berbeda karena pada prediksi tiga tahunan ini memiliki matrik data training sejumlah 93 data. Untuk data yang digunakan pada prediksi tiga tahunan adalah sebagai berikut :

Tabel 4.47 Dataset skenario tiga tahunan

Tanggal	X1	X36	Target
01/08/2010	8.479363e+09	4.489336e+09	4.806155e+09
01/09/2010	2,962,197,108	4.806155e+09	3.191364e+09
01/10/2010	3,332,520,482	3.191364e+09	4.055050e+09
.....
01/01/2021	7.554910e+09	1.316596e+10	4.055050e+09

Untuk tahapan perhitungan tidak kita tuliskan secara detail karena memiliki tahapan yang mirip dengan perhitungan pada skenario tahunan hanya jumlah data inputnya

saja yang berbeda yaitu sejumlah 36 data input sehingga memiliki matrik sebanyak 93 data training.

Untuk nilai langrange yang digunakan pada skenario tiga tahunan ini dapat dilihat pada tabel 4.48

Tabel 4.48 Nilai koefisien Langgrange

$a_1 - a_1^*$
0.420448208
0
0.420448208
.....
0.420448208

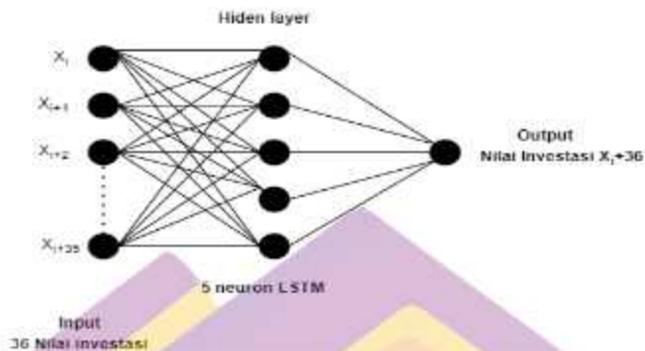
Dengan menggunakan nilai langgrange yang didapatkan dari data training akan digunakan untuk menghitung fungsi prediksi pada data testing. Hasil prediksi pada data testing adalah sebagai berikut :

Tabel 4.49 Hasil Prediksi Tiga Tahunan

Aktual	Prediksi RBF	Prediksi Polynomial
-0.94219684	-0.63275768	-0.62776167
-0.63585766	-0.63275768	-0.97594469
.....
-0.63585766	-0.63275768	-0.23759605

4.4.3.2 Hitungan manual untuk algoritma LSTM adalah sebagai berikut :

Proses pada algoritma LSTM untuk skenario prediksi tiga tahunan dengan 3 layer yaitu layer input dengan 36 data input, hidden layer dengan 5 neuron LSTM dan layer output dengan 1 data output. Arsitektur dari LSTM skenario tiga tahunan dapat dilihat pada gambar 4.11.



Gambar 4.11 Arsitektur LSTM Prediksi Tiga Tahunan

Untuk perhitungan algoritma LSTM diawali dengan membuat dataset time series tiga tahunan pada tabel 4.47 dan dilakukan normalisasi data serta pembagian data training dan data testing. Berikut adalah contoh perhitungan algoritma LSTM dengan 36 data inputan pada 1 kali iterasi untuk data tanggal 1 September 2018.

11) Menentukan nilai awal dari :

$$h_{t-1} = 0$$

$$\text{bias forget gate} = 1$$

$$\text{bias input gate} = 0.5$$

$$\text{bias input } \tilde{C}_t = 0$$

$$\text{Bias output gate} = 0.5$$

Sebelum menghitung forget gate kita menentukan nilai weight dari jumlah variabel. Untuk prediksi tiga tahunan ini memiliki variabel 36 data nilai weightnya adalah :

$$W = \left(-\frac{1}{\sqrt{36}}, \frac{1}{\sqrt{36}} \right) \\ = (-0.166, 0.166)$$

- 12) Menghitung forget gate dengan menggunakan persamaan 2.19 dan dataset yang digunakan pada tabel 4.47.

$$\begin{aligned}
 ft &= \sigma(Wf.[ht-1, xt] + bf) \\
 &= \sigma(0.166 * 0 + 0.166 * -0.473469) + 1 \\
 &= \sigma(0.9214) \\
 &= \frac{1}{1 + e^{-0.9214}} \\
 &= 0.71532
 \end{aligned}$$

- 13) Menghitung input gate dengan menggunakan dua persamaan yaitu 2.21 dan 2.22

$$\begin{aligned}
 it &= \sigma(Wi.[ht-1, xt] + bi) \\
 &= \sigma(0.166 * 0 + 0.166 * -0.473469) + 0.5 \\
 &= \sigma(0.4214) \\
 &= 0.6038
 \end{aligned}$$

$$\begin{aligned}
 \tilde{C}_t &= \tanh(Wc.[ht-1, xt] + bc) \\
 &= \tanh(0.166 * 0 + 0.166 * -0.473469) + 0 \\
 &= \tanh(-0.07859) \\
 &= 2 * \left(\frac{1}{1 + e^{-2 * (-0.07859)}} \right) - 1 \\
 &= -0.07842
 \end{aligned}$$

- 14) Menghitung update nilai cell state dengan menggunakan persamaan 2.23

$$\begin{aligned}
 Ct &= ft * C_{t-1} + it * \tilde{C}_t \\
 &= 0.71532 * 0 + 0.6038 * -0.07842 \\
 &= -0.04734
 \end{aligned}$$

15) Tahap terakhir menghitung nilai output yang menggunakan persamaan 2.24 dan 2.25

$$\begin{aligned}
 ot &= \sigma(Wo.[ht-1, xt] + bo) \\
 &= \sigma(0.166 * 0 + 0.166 * -0.473469) + 0.5 \\
 &= \sigma(0.4214) \\
 &= 0.6038 \\
 ht &= ot * \tanh(Ct) \\
 &= 0.6038 * \tanh(-0.04734) \\
 &= 0.6038 * -0.04730 \\
 &= -0.02855
 \end{aligned}$$

Pada prediksi tiga tahunan ini memiliki proses yang sama seperti pada skenario tahunan, hanya saja pada prediksi tahunan untuk nilai prediksi yang digunakan untuk dibandingkan dengan nilai aktual yang merupakan nilai inputan terakhir adalah data ke36 atau x36.

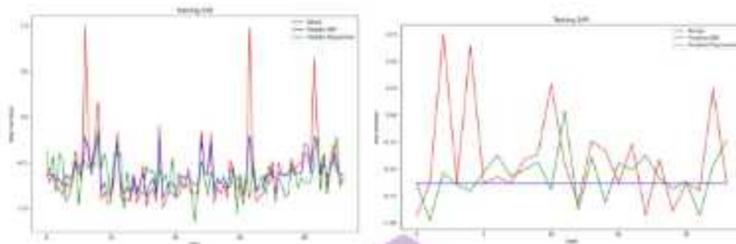
Hasil pengujian algoritma SVR dan LSTM dapat dilihat pada tabel 4.50.

Tabel 4.50 Hasil pengujian RMSE algoritma SVR & LSTM- Tiga tahun

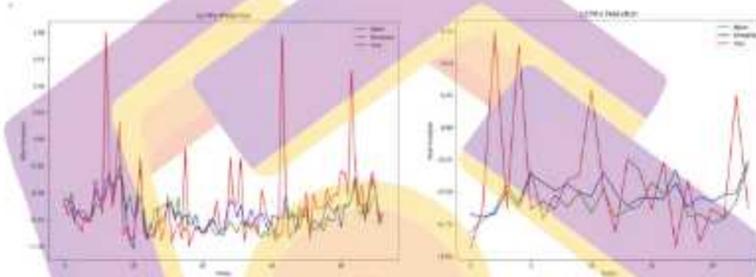
Kernel	Data Training	Data Testing
SVR - RBF	0.21916	0.50185
SVR - Polynomial	0.35485	0.51788
LSTM - Adam	0.29755	0.49599
LSTM - Rmsprop	0.33165	0.48052

Hasil pengujian pada skenario tiga tahunan menunjukkan hasil yang sama dengan prediksi bulanan dan tahunan dengan data training memiliki nilai yang lebih baik

dibandingkan data testing. Hasil prediksi tiga tahunan menggunakan algoritma SVR dengan kernel RBF memiliki hasil prediksi yang hampir sama dengan prediksi tahunan, model yang didapatkan pada data training tidak mampu diterapkan untuk data baru pada data testing sehingga hasil pengujian pada data testing menjadi tidak optimal. Hal tersebut dapat dilihat dari hasil pengujian tiga tahunan dengan nilai 0,50185 RMSE dibandingkan dengan hasil pengujian tahunan dengan nilai 0,51788 RMSE. Hasil pengujian pada algoritma LSTM menunjukkan hasil optimasi Rmsprop sedikit lebih baik dibandingkan dengan optimasi Adam dengan nilai 0.48052 RMSE. Hasil evaluasi error tersebut lebih baik dari penelitian sebelumnya yang memiliki nilai RMSE 6.119 dengan menggunakan optimasi Rmsprop dan 6.929 dengan menggunakan optimasi Adam (Soffa, 2019). Namun keduanya memiliki hasil evaluasi yang lebih baik pada optimasi Rmsprop dibandingkan optimasi Adam. Hasil pada skenario tiga tahunan ini bila dibandingkan dengan prediksi tahunan dan bulanan menjadi yang paling buruk. Hasil yang paling baik pada prediksi bulanan dengan menggunakan optimasi Rmsprop, hasil tersebut juga yang terbaik dibandingkan prediksi pada semua skenario dengan menggunakan algoritma SVR. Hasil prediksi skenario tiga tahunan dapat dilihat pada gambar 4.12 dan 4.13.



Gambar 4.12 Data Training dan data testing- SVR tiga tahun



Gambar 4.13 Data training dan data testing-LSTM tiga tahunan

Hasil pemodelan pada skenario tiga tahunan menggunakan algoritma SVR memiliki hasil dengan pola yang sama pada data training dan data testing hal tersebut karena dipengaruhi oleh jumlah data input yang banyak sejumlah 36 data input. Selain itu nilai C juga paling besar dibandingkan dengan nilai C pada prediksi bulanan dan tahunan sehingga berpengaruh pada hasil prediksi (Putu, 2020). Hasil prediksi dari ketiga skenario yaitu prediksi bulanan, tahunan dan tiga tahunan menggunakan algoritma SVR belum memperoleh hasil yang optimal. Hasil prediksi terbaik pada prediksi bulanan dengan menggunakan kernel RBF, sedangkan pada prediksi tahunan dan tiga tahunan menghasilkan prediksi yang hampir sama menggunakan kernel RBF maupun polynomial. Dari ketiga skenario hasil pemodelan skenario bulanan dengan menggunakan kernel RBF memiliki nilai

evaluasi yang paling baik dibanding skenario tahunan dan tiga tahunan. Pada algoritma LSTM untuk melakukan prediksi tiga tahunan hasil evaluasi dipengaruhi oleh jumlah data input yang sangat banyak namun jumlah data testing yang sangat sedikit. Sehingga hasil evaluasi data training memiliki hasil yang lebih baik dibandingkan dengan data testing. Hasil evaluasi dari ketiga skenario menggunakan algoritma LSTM yang terbaik pada skenario tahunan dengan optimasi Rmsprop dibandingkan pada skenario bulanan dan tiga tahunan.



BAB V

PENUTUP

5.1. Kesimpulan

Kesimpulan yang dapat diambil dari tesis ini adalah :

1. Algoritma LSTM memiliki hasil pengujian yang lebih baik dibandingkan dengan algoritma SVR karena pada hasil pengujian dengan menggunakan algoritma LSTM memiliki nilai RMSE antara 0,444 sampai 0,495, sedangkan hasil pengujian menggunakan algoritma SVR memiliki nilai RMSE antara 0,445 sampai 0,517. Namun nilai evaluasi tersebut masih dalam kategori belum optimal karena akurasi prediksi akan semakin baik apabila nilai RMSE yang semakin kecil atau mendekati nol.
2. Hasil prediksi pada algoritma SVR menunjukkan bahwa skenario bulanan memiliki hasil yang lebih baik dengan nilai 0.445 RMSE dan 0.492 RMSE, dibandingkan dengan skenario tahunan dan tiga tahunan yang memiliki hasil evaluasi lebih besar yaitu 0.501 RMSE sampai 0.517 RMSE. Sedangkan pada algoritma LSTM memiliki hasil prediksi yang lebih baik pada skenario tahunan dibandingkan prediksi bulanan maupun tiga tahunan. Nilai evaluasi pada skenario tahunan adalah 0.451 RMSE dan 0.445 RMSE sedikit lebih baik dibandingkan dengan skenario bulanan dengan nilai evaluasi 0.471 RMSE dan 0.444 RMSE maupun nilai evaluasi pada skenario tiga tahunan sebesar 0.495 RMSE dan 0.480 RMSE. Sehingga algoritma SVR cocok digunakan untuk melakukan prediksi jangka pendek

sedangkan algoritma LSTM cocok digunakan untuk melakukan prediksi jangka menengah atau tahunan untuk data realisasi investasi di Kota Magelang.

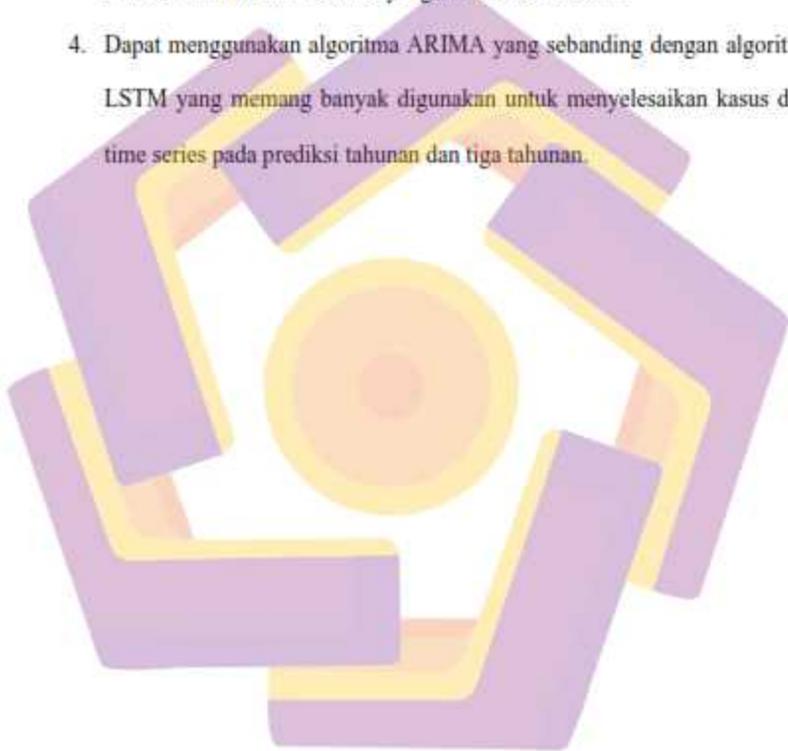
3. Algoritma LSTM lebih unggul dibandingkan dengan algoritma SVR karena nilai RMSE pada algoritma LSTM lebih kecil nilainya dibandingkan dengan nilai RMSE pada algoritma SVR seperti dapat dilihat pada nomer 1. Hal tersebut dikarenakan pada penelitian ini penentuan parameter menggunakan grid search pada algoritma SVR belum mendapatkan parameter yang optimal, selain itu dataset yang digunakan memiliki sebaran data yang mengelompok pada range tertentu sehingga model yang sudah terbentuk tidak dapat secara akurat memprediksi pada data yang baru.

5.2. Saran

Pada penelitian selanjutnya diharapkan :

1. Menggunakan data yang lebih banyak sehingga sebaran data akan lebih variatif atau tidak mengelompok pada range tertentu ataupun melakukan penanganan data outlier atau data tidak normal dengan metode yang berbeda.
2. Dapat menggunakan metode yang berbeda untuk menentukan parameter pada algoritma SVR karena pada penelitian ini menggunakan metode grid search yang dibagi dalam 2 tahap yaitu loose grid dan finer grid dan belum mendapatkan model prediksi yang baik.

3. Pada algoritma LSTM dapat menggunakan kombinasi parameter yang berbeda atau dengan melakukan pengujian parameter untuk mendapatkan model prediksi yang lebih baik. Pada penelitian ini menggunakan parameter 1 hidden layer, 5 neuron, batch size 1 dan 100 epoch namun hasil pemodelan masih memiliki nilai validasi yang belum memuaskan.
4. Dapat menggunakan algoritma ARIMA yang sebanding dengan algoritma LSTM yang memang banyak digunakan untuk menyelesaikan kasus data time series pada prediksi tahunan dan tiga tahunan.



DAFTAR PUSTAKA

PUSTAKA BUKU

Andreas Halim, 2003, Kamus Lengkap 1 Milyar Inggris-Indonesia, Sulita Jaya, Surabaya

Santosa, B, 2007, Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi, Graha Ilmu, Yogyakarta.

Prasetyo, P. Eko. 2009. Fundamental Makro Ekonomi. Yogyakarta: Beta Offset.

Vapnik, V.N, The Nature of Statistical Learning Theory, 1995, Springer, New York.

Hsu, C. W, Chang, C.C., and Lin, C.J, 2016, A Practical Guide to Support Vector Classification, Department of Computer Science National Taiwan University, Taipei.

Undang-Undang No. 25 Tahun 2007 tentang Penanaman Modal

Prof. Dr. H. Imam Ghozali, M.Com, Akt , 2011, Aplikasi Analisis Multivariate dengan Program IBM SPSS 19, Badan Penerbit Universitas Diponegoro

PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

Ivonne Margi Imanuella, Analisis Faktor yang mempengaruhi Kepercayaan Penanam Modal pada Layanan Government-To-Business dan Dampaknya Terhadap Penggunaan secara Berkelanjutan Studi Kasus Layanan Laporan Kegiatan Penanaman Modal Online, Fasikom Universitas Indonesia, 2016

Maharani, Dewi, Analisis Pengaruh Investasi dan Tenaga Kerja terhadap Produk Domestik Regional Bruto (PDRB) di Sumatera Utara Vol 8 No 2, Desember 2016

Anis Wilasih, Metode K-Means Clustering Untuk Pengalokasian Penanam Modal Asing Pada Badan Penanaman Modal Daerah Provinsi Jawa Tengah ,2015

M Abdul Dwiyanto Suyudi, Esmeralda C. Djamal, Asri Maspupah, Prediksi Harga Saham menggunakan Metode Recurrent Neural Network, Seminar Nasional Aplikasi Teknologi Informasi (SNATi), 2019

Ni Putu Ratindia Apriyanti, I Ketut Gede Darma Putra, I Made Suwija Putra, Peramalan Jumlah Kecelakaan Lalu Lintas Menggunakan Metode Support Vector Regression, Jurnal Ilmiah Merpati Vol 8 No 2, Agustus 2020

Arynda Kusuma Dewi, Prediksi Laju Pertumbuhan Penduduk menggunakan Metode Support Vector Regression (Studi Kasus : Kota Malang), Repository.ub.ac, 2020

Felix Gers, Long Short-Term Memory in Recurrent Neural Networks, Citiseer, 2001

Hendri, H. Character Recognition Dengan Menggunakan Jaringan Syaraf Tiruan. Jurnal Times Vol 32 No 2, 2014

Murtaza Roondiwala, Harshal Patel, Shraddha Varma, Predicting Stock Prices Using LSTM, International Journal of Science and Research (IJSR), 2017

Muhammad Wildan Putra Aldi, Jondri, Annisa Aditsania, Analisis Dan Implementasi Long Short Term Memory Neural Network Untuk Prediksi Harga Bitcoin, eProceedings of Engineering Vol 5 No 2, 2018

Setyowati, Ani dan Siti Fatimah, Analisis Faktor-Faktor yang Mempengaruhi Investasi Dalam Negeri di Jawa Tengah 1980-2002, Jurnal Ekonomi Pembangunan Vol. 8 No. 1, Juni 2007

Ira Melissa, Raymond S. Oetama, Analisis Data Pembayaran Kredit Nasabah Bank Menggunakan Metode Data Mining, Ultima InfoSys, Vol. IV No. 1, Juni 2013

Edwin Setya Parahita, Junaedi , 2020, Menyoal Capaian Realisasi Penanaman Modal Dalam Negeri, EcceS: Economics Social and Development Studies Vol 7 No 1, Juni 2020

Satyam Gangwar, Vikram Bali1, Ajay Kumar, Comparative Analysis of Wind Speed Forecasting Using LSTM and SVM, EAI Endorsed Transactions Vo. 7 Issue 25, 2020

D. P. Kingma and J. L. Ba, 2015, ADAM: A Method for Stochastic Optimization, ICLR, 2015.

Jannatun Khustia Lubisa , Iqbal Kharisudin, 2021, Metode Long Short Term Memory dan Generalized Autoregressive Conditional Heteroscedasticity untuk Pemodelan Data Saham ,PRISMA, Prosiding Seminar Nasional Matematika, 2021

Soffa Zahara, Sugianto, M. Bahril Ilmiddafiq, 2019, Prediksi Indeks Harga Konsumen Menggunakan Metode Long Short Term Memory (LSTM) Berbasis Cloud Computing, Jurnal Resti Vol 3 No 3, 2019

Adhib Arfan, Lussiana ETP, 2020, Perbandingan Algoritma Long Short-Term memory dengan SVR pada Prediksi Harga Saham di Indonesia Jurnal Petir Vol 13 No 1, Maret 2020

Drajat Indra Purnama , 2020, Peramalan Jumlah Penumpang Datang Melalui Transportasi Udara Di Sulawesi Tengah Menggunakan Support Vector Regression (SVR), Jurnal Ilmiah Matematika dan Terapan Vol 17 No 1, Juni 2020

Aulia Asyiva, 2019, Prediksi Laju Inflasi Indonesia menggunakan metode support Vector Regression dengan kernel Radial Basis Function, UIN Syarif Hidayatullah Jakarta, 2019

Rya Sofi Aulia dan Raden Mohamad Atok, Penentuan Panjang Optimal Data Deret Waktu Bebas Outlier dengan Menggunakan Metode Window Time, JURNAL SAINS DAN SENI ITS Vol. 6, No. 1, 2017

Ekky Novrizal Alam, Prediksi Kebutuhan Obat dengan Support Vector Regression dan Long Short Term Memory, Institute Teknologi Bandung, 2019

Wulan Wahyu Kusuma Wardani, Prediksi Harga Saham Syariah Menggunakan Metode Recurrent Neural Network-Long Short Term Memory, Universitas Islam Negeri Sunan Ampel Surabaya, 2021

I Gusti Ayu Made Srinadi, Pengaruh Outlier terhadap Estimator Parameter Regresi dan metode Regresi Robust, Institut Teknologi Sepuluh November, 2014

PUSTAKA LAPORAN PENELITIAN

Indarto, 2020, Perbandingan Prediksi Resiko Kematian Pasien Stroke Perdarahan menggunakan Algoritma Decision Tree C4.5 dan Naive Bayes sebelum dan sesudah seleksi atribut, Tesis, Magister Teknik Informatika, STIMIK AMIKOM, Yogyakarta

