

TESIS

**PREDIKSI INDEKS HARGA KONSUMEN KOMODITAS MAKANAN DI
JAWA TIMUR MENGGUNAKAN SUPPORT VECTOR REGRESSION
DAN POLYNOMIAL REGRESSION**



Disusun oleh:

Nama : Ayu Adellna Suyono
NIM : 20.55.1344
Konsentrasi : Business Intelligence

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

2023

TESIS

**PREDIKSI INDEKS HARGA KONSUMEN KOMODITAS MAKANAN DI
JAWA TIMUR MENGGUNAKAN SUPPORT VECTOR REGRESSION
DAN POLYNOMIAL REGRESSION**

**PREDICTION OF CONSUMER PRICE INDEX FOR FOOD
COMMODITIES IN EAST JAVA USING SUPPORT VECTOR
REGRESSION AND POLYNOMIAL REGRESSION**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Ayu Adellna Suyono
NIM : 20.55.1344
Konsentrasi : Business Intelligence

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

2023

HALAMAN PENGESAHAN

**PREDIKSI INDEKS HARGA KONSUMEN KOMODITAS MAKANAN DI JAWA
TIMUR MENGGUNAKAN SUPPORT VECTOR REGRESSION DAN
POLYNOMIAL REGRESSION**

**PREDICTION OF CONSUMER PRICE INDEX FOR FOOD COMMODITIES IN
EAST JAVA USING SUPPORT VECTOR REGRESSION AND POLYNOMIAL
REGRESSION**

Dipersiapkan dan Disusun oleh

Ayu Adellna Suyono

20.55.1344

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Rabu, 4 Januari 2023

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

Yogyakarta, 4 Januari 2023

Rektor

Prof. Dr. M. Suyanto, M.M.

NIK. 190302001

HALAMAN PERSETUJUAN

PREDIKSI INDEKS HARGA KONSUMEN KOMODITAS MAKANAN DI JAWA TIMUR MENGGUNAKAN SUPPORT VECTOR REGRESSION DAN POLYNOMIAL REGRESSION

PREDICTION OF CONSUMER PRICE INDEX FOR FOOD COMMODITIES IN EAST JAVA USING SUPPORT VECTOR REGRESSION AND POLYNOMIAL REGRESSION

Dipersiapkan dan Disusun oleh

Ayu Adelina Suyono

20.55.1344

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Rabu, 4 Januari 2023

Pembimbing Utama

Anggota Tim Penguji

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

Dhani Arlatmanto, M.Kom., Ph.D.
NIK. 190302197

Pembimbing Pendamping

Dr. Kumara Ari Yuana, S.T., M.T.
NIK. 190302575

M. Rudvanto Arief, M.T.
NIK. 190302098

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

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

Yogyakarta, 4 Januari 2023
Direktur Program Pascasarjana

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

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ayu Adelfina Suyono
NIM : 20.55.1344
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul berikut:
Prediksi Indeks Harga Konsumen Komoditas Makanan di Jawa Timur menggunakan Support Vector Regression dan Polynomial Regression

Dosen Pembimbing Utama : Prof. Dr. Kusriani, M.Kom
Dosen Pembimbing Pendamping : M. Rudyanto Arief, M.T.

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, 4 Januari 2023

Yang Menyatakan,


METERA
TEMPAT
MAKX207021603

Ayu Adelfina Suyono

HALAMAN PERSEMBAHAN

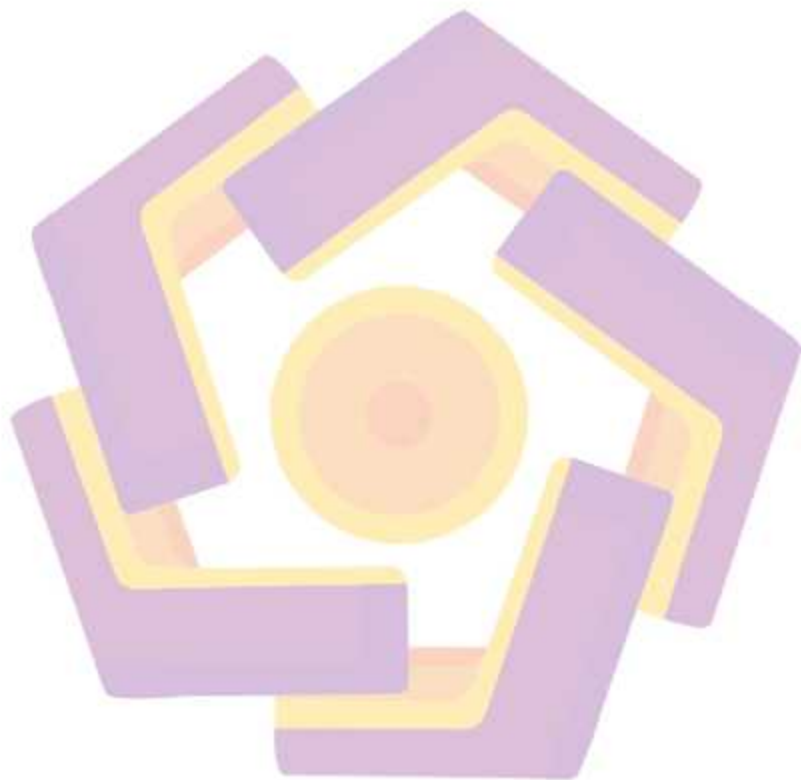
Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat-Nya, sehingga penulis bisa menyelesaikan penelitian tesis tepat pada waktunya. Shalawat serta salam tak lupa dihaturkan untuk Nabi Muhammad SAW yang telah menuntun manusia untuk senantiasa menggali ilmu Allah SWT, sebagaimana tertuang dalam hadits, *Uthlubul 'Ilma minal Mahdi ilal Lahdi* yang artinya *Tuntutlah Ilmu dari Buaian hingga Liang Lahat*.

Selain itu penulis juga ingin mempersembahkan penelitian tesis ini kepada pihak-pihak yang telah mendukung, antara lain:

1. Ayah (Alm. Masherri Suyono), Ibu (Ma'rifah Izah), dan Adik (Diah Ajeng Valina Suyono) yang telah memberikan kasih sayang yang tidak terkira;
2. Suami (Shobirin Nugroho) dan Anak (Zayba Almira Nugroho) atas kasih sayang dan dukungan yang diberikan,;
3. Sahabat, rekan, dan kolega yang turut memberikan do'a dan dukungannya dalam proses penelitian tesis ini;
4. Kepada semua pihak yang membantu pengerjaan penelitian tesis ini, semoga tesis ini dapat memberikan manfaat bagi semua.

HALAMAN MOTTO

"Practice Makes Perfect – Julie James"



KATA PENGANTAR

Puji syukur kehadiran Allah SWT Yang Maha Pengasih dan Penyayang, sehingga penulis dapat menyelesaikan penelitian tesis dengan judul “Prediksi Indeks Harga Konsumen Komoditas Makanan di Jawa Timur menggunakan Support Vector Regression dan Polynomial Regression”. Tesis ini tersusun sebagai salah satu syarat untuk memperoleh gelar Magister Komputer di Universitas Amikom Yogyakarta.

Penulis menyadari bahwa dalam penyusunan tesis ini masih memiliki kekurangan. Maka dari itu penulis mengharapkan dan menerima saran serta kritik demi perbaikan dan kesempurnaan penulisan tesis ini. Tesis ini dapat terselesaikan tidak terlepas dari doa, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan rasa terima kasih kepada:

1. Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta;
2. Prof. Dr. Kusrini, M.Kom. selaku Direktur Program Pascasarjana Universitas Amikom Yogyakarta, Ketua Program Studi S2 PJJ Teknik Informatika, sekaligus Pembimbing Utama;
3. Prof. Dr. Ema Utami, S.Si., M.Kom. selaku Wakil Direktur Program Pascasarjana Universitas Amikom Yogyakarta;
4. Alva Hendi Muhammad, M.Eng., Ph.D. selaku Sekretaris Program Studi S2 PJJ Teknik Informatika;
5. M. Rudyanto Arief, M.T. selaku Pembimbing Kedua;
6. Dhani Ariatmanto, M.Kom., Ph.D. selaku Penguji 1;

7. Dr. Kumara Ari Yuana, S.T., M.T. selaku Penguji 2;
8. Prof. Dr. Kusriani, M.Kom. selaku Penguji 3;
9. Segenap dewan dosen dan tutor Program Jarak Jauh Magister Teknik Informatika Universitas Amikom Yogyakarta;
10. Seluruh staf pengelola Pascasarjana Universitas Amikom Yogyakarta;

Serta seluruh pihak yang berperan dalam penelitian tesis ini dan tidak bisa penulis sebutkan satu per satu. Semoga segala do'a, bimbingan, dan dukungan yang diberikan akan digantikan oleh Allah SWT pahala yang berlipat. Akhir kata penulis berharap penelitian tesis ini dapat bermanfaat bagi penulis pada khususnya dan bagi pembaca pada umumnya. Aamiin.

Yogyakarta, 4 Januari 2023

Penulis

DAFTAR ISI

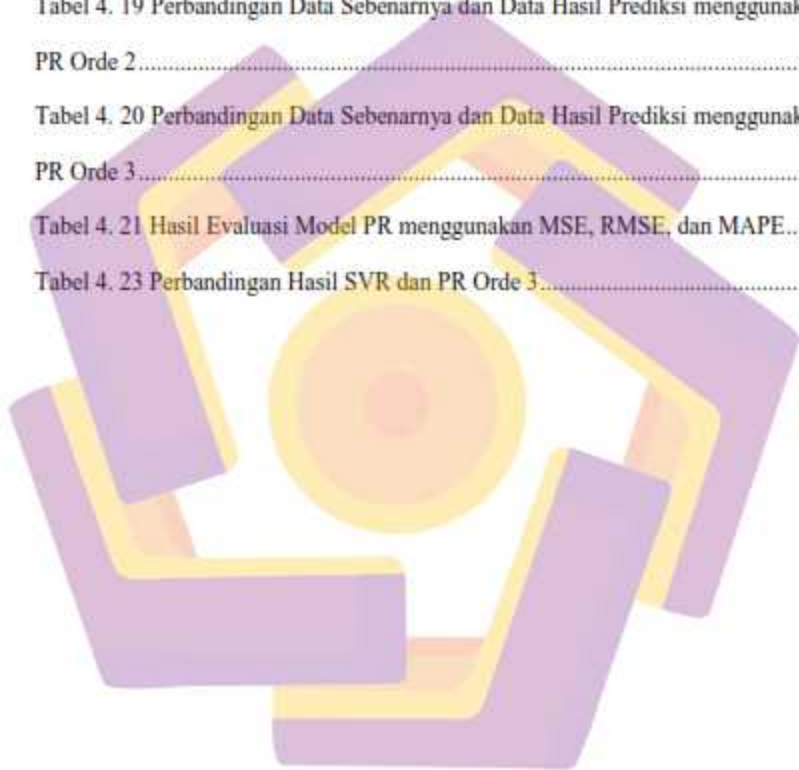
HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS.....	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiv
INTISARI.....	xvii
<i>ABSTRACT</i>	xviii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	6
1.3. Batasan Masalah.....	6
1.4. Tujuan Penelitian.....	8
1.5. Manfaat Penelitian.....	8
BAB II TINJAUAN PUSTAKA.....	10
2.1. Tinjauan Pustaka.....	10
2.2. Keaslian Penelitian.....	13

2.3. Landasan Teori.....	18
2.3.6. Evaluasi Hasil Prediksi	32
BAB III METODE PENELITIAN.....	36
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	36
3.2. Metode Pengumpulan Data.....	37
3.3. Metode Analisis Data.....	37
3.4. Alur Penelitian.....	38
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	42
4.1. Perencanaan Aksi.....	42
4.1.1. Pengumpulan Data.....	42
4.1.2. Praproses Data.....	51
4.1.3. Pemisahan Data menjadi Data Latih dan Data Uji.....	52
4.2. Pelaksanaan Aksi.....	54
4.2.1. Analisa Korelasi menggunakan Pearson Product Moments.....	54
4.2.2. Prediksi Nilai IHK menggunakan Support Vector Regression.....	60
4.2.3. Prediksi Nilai IHK menggunakan Polynomial Regression.....	70
4.2.4. Perbandingan Prediksi IHK menggunakan SVR dan PR.....	77
BAB V PENUTUP.....	81
5.1. Kesimpulan.....	81
5.2. Saran.....	82
DAFTAR PUSTAKA.....	84

DAFTAR TABEL

Tabel 2. 1 Matriks literatur review dan posisi penelitian Prediksi Indeks Harga Konsumen Komoditas Makanan di Jawa Timur menggunakan Support Vector Regression dan Polynomial Regression.....	13
Tabel 2. 2 Kriteria Nilai MAPE.....	35
Tabel 4. 1 Contoh Data Harian Komoditas Makanan.....	43
Tabel 4. 2 Data IHK Kabupaten Jember.....	44
Tabel 4. 3 Data IHK Kabupaten Banyuwangi.....	45
Tabel 4. 4 Data IHK Kabupaten Sumenep.....	45
Tabel 4. 5 Data IHK Kota Kediri.....	46
Tabel 4. 6 Data IHK Kota Malang.....	46
Tabel 4. 7 Data IHK Kota Probolinggo.....	46
Tabel 4. 8 Data IHK Kota Madiun.....	47
Tabel 4. 9 Data IHK Kota Surabaya.....	47
Tabel 4. 10 Tabel Variabel Penentu Nilai IHK.....	48
Tabel 4. 11 Contoh Data Hasil Integrasi Harga Komoditas Makanan dan IHK... 52	52
Tabel 4. 12 Proporsi Perbandingan Data Latih dan Data Uji.....	53
Tabel 4. 13 Data Hasil Korelasi.....	57
Tabel 4. 14 Tingkat Korelasi Variabel X dan Variabel Y.....	59
Tabel 4. 15 Hasil Hyperparameter Tuning menggunakan Grid Search Optimization untuk semua Perbandingan Data.....	63
Tabel 4. 16 Hasil Evaluasi SVR dengan Perbandingan Data yang Berbeda.....	66

Tabel 4. 17 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan SVR.....	68
Tabel 4. 18 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 1.....	71
Tabel 4. 19 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 2.....	73
Tabel 4. 20 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 3.....	75
Tabel 4. 21 Hasil Evaluasi Model PR menggunakan MSE, RMSE, dan MAPE..	76
Tabel 4. 23 Perbandingan Hasil SVR dan PR Orde 3.....	78



DAFTAR GAMBAR

Gambar 2. 1 Pengaturan Soft Margin Loss untuk Linear SVM.....	24
Gambar 2. 2 Diagram Pencar Korelasi Pearson Product Moments	32
Gambar 3. 1 Alur Penelitian.....	39
Gambar 4. 1 Impor Packages Python.....	50
Gambar 4. 2 Impor Dataset.....	50
Gambar 4. 3 Hasil Data yang Sudah Diimpor.....	50
Gambar 4. 4 Hasil Info Data yang Sudah Diimpor.....	51
Gambar 4. 5 Proses Penghapusan Missing Value.....	52
Gambar 4. 6 Import Packages.....	54
Gambar 4. 7 Import Dataset.....	54
Gambar 4. 8 Proses Penentuan Variabel X dan Variabel Y.....	55
Gambar 4. 9 Plot Diagram Pencar.....	56
Gambar 4. 10 Visualisasi Diagram Pencar Analisa Korelasi dengan Pearson Product Moment.....	56
Gambar 4. 11 Proses Penampilan Data Hasil Korelasi.....	57
Gambar 4. 12 Proses Pencarian Hyperparameter menggunakan Grid Search Optimization.....	61
Gambar 4. 13 Hasil Hyperparameter Tuning menggunakan Grid Search Optimization untuk Perbandingan 70%:30%.....	61
Gambar 4. 14 Hasil Hyperparameter Tuning menggunakan Grid Search Optimization untuk Perbandingan 80%:20%.....	62

Gambar 4. 15 Hasil Hyperparameter Tuning menggunakan Grid Search Optimization untuk Perbandingan 90%:10%.....	63
Gambar 4. 16 Implementasi Nilai Hyperparameter	64
Gambar 4. 17 Implementasi Model SVR berdasarkan Data Latih dengan perbandingan 90%:10%	64
Gambar 4. 18 Implementasi Model SVR berdasarkan Data Uji dengan perbandingan 90%:10%	64
Gambar 4. 19 Implementasi Model SVR berdasarkan Data Latih dengan perbandingan 80%:20%	65
Gambar 4. 20 Implementasi Model SVR berdasarkan Data Uji dengan perbandingan 80%:20%	65
Gambar 4. 21 Implementasi Model SVR berdasarkan Data Latih dengan perbandingan 70%:30%	65
Gambar 4. 22 Implementasi Model SVR berdasarkan Data Uji dengan perbandingan 70%:30%	65
Gambar 4. 23 Plot Data Sebenarnya dan Data Hasil Prediksi menggunakan SVR untuk Perbandingan Data Latih dan Data Uji 90%:10%	67
Gambar 4. 24 Plot Data Sebenarnya dan Data Hasil Prediksi menggunakan SVR untuk Perbandingan Data Latih dan Data Uji 80%:20%	67
Gambar 4. 25 Plot Data Sebenarnya dan Data Hasil Prediksi menggunakan SVR untuk Perbandingan Data Latih dan Data Uji 70%:30%	68
Gambar 4. 26 Implementasi Model PR Orde 1	70
Gambar 4. 27 Implementasi Model PR Orde 1 berdasarkan Data Latih	70

Gambar 4. 28 Implementasi Model PR Orde 1 berdasarkan Data Uji.....	71
Gambar 4. 29 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 1	71
Gambar 4. 30 Implementasi Model PR Orde 2.....	72
Gambar 4. 31 Implementasi Model PR Orde 2 berdasarkan Data Latih	72
Gambar 4. 32 Implementasi Model PR Orde 2 berdasarkan Data Uji.....	72
Gambar 4. 33 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 2	73
Gambar 4. 34 Implementasi Model PR Orde 3	74
Gambar 4. 35 Implementasi Model PR Orde 3 berdasarkan Data Latih	74
Gambar 4. 36 Implementasi Model PR Orde 3 berdasarkan Data Uji.....	74
Gambar 4. 37 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 3	75
Gambar 4. 38 Perbandingan Plot Data Sebenarnya dan Data Hasil Prediksi SVR dan PR Orde 3	79

INTISARI

Makanan merupakan salah satu kebutuhan pokok yang wajib dipenuhi oleh manusia. Tidak peduli naik atau turun harga komoditas makanan, kebutuhan akan makanan tidak bisa digantikan. Setelah diguncang oleh pandemi Covid-19 mulai awal tahun 2020 lalu, Indonesia mengalami pasang surut dalam tingkat perekonomiannya. Beberapa komoditas menjadi langka dan banyak yang mengalami kenaikan harga. Kenaikan harga komoditas menimbulkan turunnya daya beli masyarakat yang dapat pula berpengaruh pada perekonomian negara.

Indeks Harga Konsumen (IHK) merupakan suatu nilai yang terbentuk hasil dari pengolahan data komoditas. Nilai ini dapat dijadikan sebagai salah satu acuan dalam menentukan adanya suatu inflasi atau deflasi ada suatu negara. Pada penelitian ini akan memanfaatkan *machine learning* untuk proses prediksi nilai IHK menggunakan data harga komoditas makanan.

Data harga komoditas makanan dan data IHK yang digunakan berasal dari 34 komoditas makanan yang umum dikonsumsi di Jawa Timur. Periode data yang digunakan adalah pada tahun 2014-2020 dan terdiri dari 8 kabupaten/kota. Data tersebut kemudian diolah untuk dicari adanya korelasi antar variabel dan prediksi nilai IHK-nya. Analisa korelasi dilakukan dengan Pearson Product Moment, sedangkan proses prediksi dilakukan dengan 2 model. Model yang digunakan adalah Support Vector Regression dan Polynomial Regression.

Hasil Analisa korelasi menunjukkan adanya korelasi negatif sempurna antara variabel x (harga komoditas makanan) dan variabel y (nilai IHK). Artinya semakin naik harga komoditas makanan, maka semakin turun nilai IHK-nya. Sedangkan hasil prediksi menggunakan model SVR menghasilkan nilai MAPE 5,9% untuk data latih dan 6% untuk data uji. Sedangkan model PR menghasilkan nilai MAPE 0,3% pada data latih dan 3,4% pada data uji (Orde 3). Berdasarkan hasil perbandingan nilai MAPE dapat disimpulkan bahwa metode dengan performa terbaik adalah PR dengan Orde 3.

Kata kunci: komoditas makanan, ihk, analisa korelasi, prediksi, pearson product moment, support vector regression, polynomial regression.

ABSTRACT

Food is one of the basic needs that must be met by humans. No matter the price of food commodities rises or falls, the need for food cannot be replaced. After being rocked by the Covid-19 pandemic starting in early 2020, Indonesia has experienced ups and downs in its economic level. Some commodities are becoming scarce and many are experiencing price increases. The increase in commodity prices causes a decrease in people's purchasing power which can also affect the country's economy.

The Consumer Price Index (CPI) is a value formed as a result of processing commodity data. This value can be used as a reference in determining the existence of an inflation or deflation in a country. In this study, machine learning will be used to predict the CPI value using food commodity price data.

Food commodity price data and CPI data used are derived from 34 food commodities that are commonly consumed in East Java. The data period used is in 2014-2020 and consists of 8 districts/cities. The data is then processed to look for correlations between variables and the prediction of the CPI value. Correlation analysis was carried out with Pearson Product Moment, while the prediction process was carried out with 2 models. The models used are Support Vector Regression and Polynomial Regression.

The results of the correlation analysis showed that there was a perfect negative correlation between the x variable (food commodity prices) and the y variable (CPI value). This means that the higher the price of food commodities, the lower the CPI value. While the prediction results using the SVR model produce a MAPE value of 5.9% for training data and 6% for test data. While the PR model produces a MAPE value of 0.3% in the training data and 3.4% in the test data (Orde 3). Based on the results of the comparison of MAPE values, it can be concluded that the method with the best performance is PR with Orde 3.

Keyword: food commodities, CPI, correlation analysis, prediction, pearson product moment, support vector regression, polynomial regression

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Harga komoditas makanan di Indonesia merupakan salah satu data yang memiliki sifat fluktuatif. Artinya data ini kerap kali mengalami kenaikan dan penurunan. Skala perubahan yang terjadi pun terbilang cepat. Tidak hanya dalam skala bulanan ataupun mingguan, namun juga dalam skala harian.

Beberapa waktu yang lalu, Indonesia sempat mengalami fenomena kenaikan harga minyak goreng yang cukup drastis. Harga minyak goreng yang awalnya berada pada kisaran harga Rp 10.000,- hingga Rp 15.000,- per liter, melejit di angka Rp 20.000,- hingga Rp 30.000,- per liter. Pada dasarnya tidak hanya minyak goreng saja yang memiliki tren harga yang fluktuatif, melainkan komoditas makanan lain seperti cabai, bawang, telur, dan komoditas lainnya juga memiliki tren harga yang naik dan turun.

Perubahan harga komoditas, terutama ketika terjadi kenaikan harga pasti akan berdampak pada perekonomian negara. Ketika terjadi kenaikan harga, daya beli masyarakat pada suatu komoditas akan mengalami penurunan. Jika hal tersebut terjadi, maka pendapatan negara pun akan turut mengalami penurunan.

Pemerintah dalam hal ini Badan Pusat Statistik (BPS) memiliki suatu metode yang dapat digunakan untuk mengetahui adanya kenaikan dan penurunan tingkat perekonomian dalam negara. Salah satu metode yang digunakan adalah dengan memanfaatkan Indeks Harga Konsumen (IHK).

IHK merupakan suatu indeks atau nilai yang diperoleh dari hasil pengolahan data harga komoditas makanan dalam suatu kurun waktu tertentu. IHK dapat digunakan oleh pemerintah sebagai salah satu metode untuk mengukur adanya inflasi atau deflasi dalam perekonomian sehingga pemerintah dapat memberikan langkah yang tepat untuk mengatasi persoalan yang terjadi.

Pada penelitian ini akan dirancang dan dilakukan eksperimen untuk prediksi nilai IHK menggunakan metode dari mesin pembelajaran. Pada kasus prediksi IHK, metode yang digunakan adalah mesin pembelajaran dalam fokus regresi. Regresi merupakan suatu metode yang digunakan untuk memperkirakan hubungan yang terjadi antara variabel terikat dan variabel bebas. Dalam penelitian ini metode regresi digunakan untuk memperkirakan hubungan antara nilai IHK sebagai variabel terikat dengan harga komoditas makanan sebagai variabel bebasnya.

Untuk melakukan sebuah proses regresi, variabel-variabel yang terlibat harus memiliki korelasi. Hal ini penting dilakukan, karena jika dalam proses regresi dilibatkan variabel yang tidak memiliki korelasi, maka hasil regresi tidak akan maksimal. Variabel dengan korelasi yang tinggi atau kuat biasanya akan menghasilkan prediksi yang akurat, begitu pula sebaliknya. Namun tidak menutup kemungkinan pula korelasi rendah memiliki akurasi yang tinggi begitu pula sebaliknya. Dalam penelitian ini akan digunakan dua macam variabel, yakni harga komoditas yang terdiri dari 34 komoditas makanan dan juga nilai IHK. Analisa korelasi akan dilakukan menggunakan metode Pearson Product Moments. Jika masing-masing variabel tersebut memiliki korelasi, maka berarti variabel tersebut dapat digunakan untuk proses regresi.

Terdapat beberapa macam metode yang digunakan dalam kasus regresi. Beberapa metode yang terkenal, diantaranya Linear Regression, Polynomial Regression, Bayesian Linear Regression, Support Vector Regression, dan lain sebagainya. Semua metode yang telah disebutkan tentu memiliki kelebihan dan kekurangan masing-masing sesuai karakteristik data yang digunakan. Dalam penelitian ini akan digunakan dua jenis metode, yaitu Support Vector Regression (SVR) dan Polynomial Regression (PR).

SVR merupakan metode yang berasal dari Support Vector Machine (SVM). Metode ini dikenal juga sebagai turunan dari SVM yang digunakan untuk mengatasi masalah regresi. Algoritma ini dikenal mampu mengatasi *overfitting* data dengan baik sehingga dapat menghasilkan prediksi dengan tingkat kesalahan yang rendah. SVR cocok digunakan untuk data yang bersifat kontinyu, sehingga sesuai dengan karakteristik data harga komoditas makanan dan nilai IHK. SVR juga dikenal mampu mengatasi data dengan dimensi yang tinggi dengan bantuan dari fungsi kernel. Seperti yang diketahui, bahwa data harga komoditas makanan merupakan data harian dan data yang berdimensi tinggi, maka SVR merupakan metode yang cocok untuk digunakan dalam proses prediksi IHK.

Beberapa penelitian menunjukkan keandalan SVR dalam melakukan regresi. Penelitian pertama dilakukan untuk melakukan estimasi status kesehatan baterai lithium-ion (Li *et al.*, 2022). Metode SVR digunakan untuk melakukan estimasi terkomputerisasi dengan harapan akan menghasilkan estimasi yang lebih akurat daripada metode tradisional. Proses pemodelan dimulai dengan melakukan analisis korelasi antara empat fitur yang berhubungan dengan kesehatan baterai.

Analisa korelasi Pearson dipilih sebagai metode yang digunakan. Hasil analisa korelasi menunjukkan korelasi yang tinggi antara empat fitur yang dimaksud dengan status kesehatan baterai. Dilanjutkan dengan proses estimasi dengan metode SVR. Hasil yang ditunjukkan memiliki performa yang bagus, dengan tingkat kesalahan kurang dari 2%.

Penelitian SVR yang selanjutnya dilakukan untuk melakukan prediksi pada timbunan sampah tahunan di Kota Bahrain (Jassim, Coskuner and Zontul, 2022). Pada penelitian ini digunakan beberapa kernel dari SVR, seperti Kernel Linear, Polinomial, dan Radial Basis Function (RBF). Hasil prediksi terbaik ditunjukkan oleh prediksi menggunakan Kernel RBF dengan nilai R^2 sebesar 0,97% dan 4,82% untuk pengujian dengan data latih dan data uji. Berdasarkan hasil dari penelitian yang telah dipaparkan dapat ditarik kesimpulan bahwa metode SVR data digunakan untuk proses prediksi dengan tingkat kesalahan yang cukup rendah. Tingkat kesalahan yang dihasilkan kurang dari 10%, yang menurut kriteria maka model regresi dapat dikatakan memiliki performa sangat baik.

Metode kedua yang akan digunakan adalah Polynomial Regression. Metode ini merupakan kelanjutan dari Regresi Linear Berganda (RLB). Perbedaan yang dimiliki adalah bahwa pada PR perlu ditambahkan adanya Orde. Orde ini dimulai dengan Orde 1. Jika yang digunakan adalah Orde 1, maka formula yang digunakan akan sama persis seperti RLB. Jika Orde yang digunakan adalah Orde 2, maka formula yang digunakan sama dengan RLB dipangkatkan dengan 2. Jika Orde 3 akan dipangkatkan dengan 3, dan seterusnya. Semakin tinggi orde yang digunakan,

maka model yang dihasilkan akan semakin kompleks. Namun hal ini tidak bisa menjadi jaminan bahwa model yang dihasilkan akan semakin baik.

Beberapa penelitian dilakukan dengan memanfaatkan PR. Penelitian pertama dilakukan dalam penelitian tesis yang dilakukan oleh (Erfiana, 2015). Penelitian ini dilakukan untuk melakukan prediksi perubahan nilai tanah di wilayah Eksplorasi Minyak Gas dan Bumi Bojonegoro. Pada penelitian yang dilakukan oleh Erfiana, ternyata hasil yang ditunjukkan oleh PR memiliki performa yang kurang baik. Tingkat kesalahan yang ditunjukkan mencapai 36,6%. Berdasarkan pernyataan yang dituliskan oleh penulis disebutkan bahwa metode ini kurang cocok diterapkan di wilayah penelitian.

Penelitian kedua dilakukan oleh (Eka *et al.*, 2021). Pada penelitian ini Eka, dkk. menggunakan PR untuk melakukan prediksi produksi daging sapi nasional. Penelitian ini membandingkan metode regresi linear dan regresi polinomial. Berdasarkan hasil dari penelitian ini mendapatkan kesimpulan bahwa hasil prediksi menggunakan regresi linear lebih unggul dengan nilai R^2 sebesar 0,9115. Sementara PR Orde 4 menghasilkan nilai R^2 sebesar 0,9181. Meskipun regresi linear lebih unggul namun performa dari PR juga cenderung cukup baik.

Berbeda dengan hasil yang diberikan oleh SVR, nampaknya PR memiliki performa yang kurang memuaskan. Padahal data yang digunakan dalam penelitian sudah memenuhi kriteria yang dibutuhkan, yaitu data yang bersifat kontinyu. Kekurangan dalam PR salah satunya belum dapat mengatasi *overfitting* data dengan baik. Itu artinya model ini bisa jadi memberikan hasil model yang baik pada data

pelatihannya, namun tidak untuk data yang baru. Sementara pada SVR, model sudah dapat mengatasi *overfitting* data dengan baik.

Berdasarkan pemaparan tersebut dapat pula disimpulkan bahwa SVR dan PR sama-sama merupakan model yang digunakan untuk proses regresi. Namun kedua model memiliki sifat yang bertolak belakang. Maka dari itu dalam penelitian ini penulis akan melakukan sebuah eksperimen untuk membandingkan apakah kedua model yang bertolak belakang ini dapat diimplementasikan dalam proses prediksi IHK. Kedua model akan diimplementasikan dalam proses prediksi IHK. Hasil yang didapatkan kemudian akan dibandingkan untuk mencari metode dengan performa terbaik berdasarkan hasil evaluasi.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, maka rumusan masalah dalam penelitian ini antara lain:

1. Bagaimana korelasi antara variabel penentu nilai IHK (harga komoditas makanan) terhadap nilai IHK?
2. Bagaimana hasil perbandingan tingkat kesalahan algoritma SVR dan PR untuk prediksi IHK komoditas makanan di Jawa Timur?

1.3. Batasan Masalah

Beberapa hal yang menjadi fokus dalam penelitian ini akan dibahas pada sub bab Batasan masalah. Batasan masalah dalam penelitian ini antara lain:

1. Data yang digunakan pada penelitian ini merupakan data harga konsumen komoditas bahan makanan di 8 kabupaten/kota di Jawa Timur, yaitu Kabupaten

Jember, Kabupaten Banyuwangi, Kabupaten Sumenep, Kota Kediri, Kota Malang, Kota Probolinggo, Kota Madiun, dan Kota Surabaya;

2. Data tersebut terdiri dari 34 harga komoditas bahan pangan pokok, diantaranya Beras Bengawan, Beras Mentik, Beras IR64, Gula Pasir Dalam Negeri, Minyak Goreng Bimoli Botol/Kemasan 2 ltr, Minyak Curah, Daging Sapi Murni, Daging Ayam Broiler, Daging Ayam Kampung, Telur Ayam Ras/Petelur, Telur Ayam Kampung, Susu Kental Manis Merk Bendera, Susu Kental Manis Merk Indomilk, Susu Bubuk Merk Bendera (Instant), Susu Bubuk Merk Indomilk (Instant), Jagung Pipilan Kering, Garam yodium jenis Bata, Garam yodium jenis Halus, Tepung Terigu Segitiga Biru (Kw Medium), Kacang Kedelai Eks Impor, Mie Instant Indomie Rasa Kari Ayam, Cabe Biasa, Cabe Rawit, Bawang Merah, Bawang Putih, Ikan Asin Teri, Kacang Hijau, Kacang Tanah, Ketela Pohon, Kol/Kubis, Kentang, Tomat, Wortel, dan Buncis;
3. Data pada poin 1 dan 2 sebagai data masukan diambil dari <https://www.siskaperbapo.jatimprov.go.id/>;
4. Data IHK Provinsi Jawa Timur sebagai data keluaran diambil dari <https://jatim.bps.go.id/>;
5. Periode data yang digunakan berasal dari tahun 2014-2020;
6. Data masukan dan data keluaran tersebut akan digunakan untuk analisis korelasi variabel dan melakukan prediksi IHK;
7. Analisis korelasi variabel untuk mengetahui pengaruh variabel masukan harga komoditas makanan terhadap variabel keluaran nilai IHK menggunakan Pearson Product Moment;

8. Prediksi IHK menggunakan algoritma SVR dan PR;
9. Skenario evaluasi dari hasil prediksi IHK komoditas makanan di Jawa Timur menggunakan Mean Square Error (MSE), Root Mean Square Error (RMSE), dan Mean Absolute Percentage Error (MAPE);
10. Alat yang digunakan dalam proses implementasi adalah Jupyter Notebook dengan bahasa pemrograman Python;
11. Hasil penelitian akan disajikan dalam bentuk model perhitungan dari korelasi variabel dan hasil prediksi menggunakan algoritma SVR dan PR.

1.4. Tujuan Penelitian

Tujuan diadakannya penelitian ini antara lain:

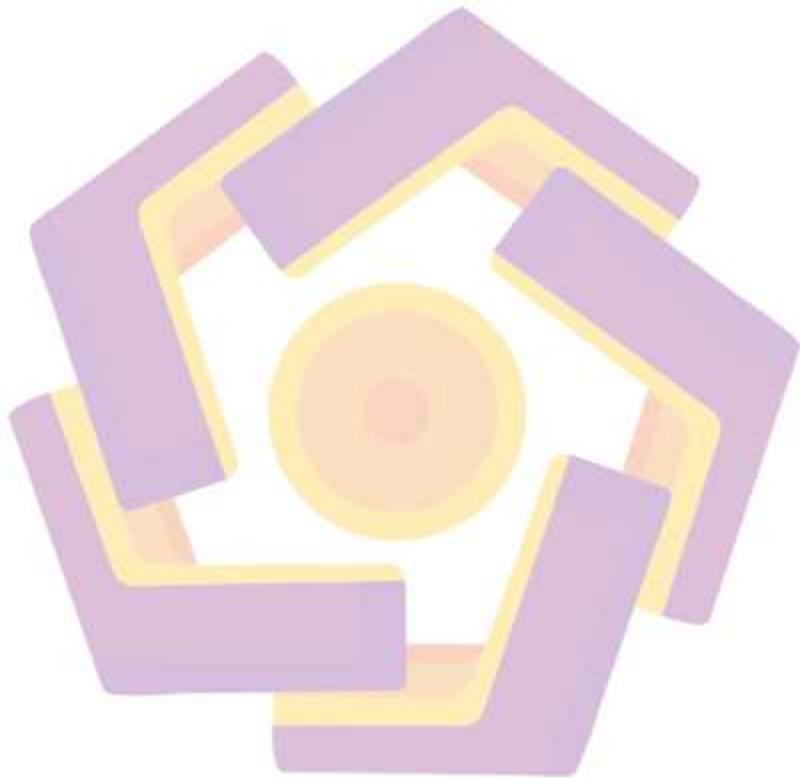
1. Mengetahui korelasi variabel penentu nilai IHK (harga komoditas makanan) terhadap nilai IHK;
2. Mengetahui hasil perbandingan tingkat kesalahan algoritma SVR dan PR untuk prediksi IHK komoditas makanan di Jawa Timur.

1.5. Manfaat Penelitian

Penelitian ini diharapkan mampu memberikan manfaat dan sumbangsih yang berguna bagi dunia ilmu pengetahuan. Manfaat penelitian antara lain:

1. Memberikan rekomendasi metode untuk prediksi IHK komoditas makanan di Jawa Timur;
2. Mengetahui pengaruh variabel masukan terhadap variabel keluaran dalam penentuan nilai IHK;

3. Memberikan kontribusi pada kebaruan penelitian mengenai prediksi IHK komoditas makanan di Jawa Timur.



BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Beberapa penelitian pernah dilakukan untuk melakukan prediksi IHK. Metode kembar *time series* yaitu ARIMA, SARIMA, dan VARIMA menjadi algoritma yang cukup sering digunakan (Ananda, Tarno and Sudarno, 2020; Ayudhiah, Bahri and Fitriyani, 2020; Dimashanti and Sugiman, 2021; Hali Mukron *et al.*, 2021). Pada penelitian Ananda, dkk menggunakan ARIMA sebagai metode analisis data. Setelah proses analisis data dilakukan kemudian dilanjutkan dengan identifikasi model fungsi transfer. Hasil implementasi ARIMA Bersama dengan model fungsi transfer input menghasilkan nilai MAPE yang rendah yaitu pada 2,351591% (Ananda, Tarno and Sudarno, 2020). Penelitian selanjutnya dilakukan menggunakan Vector Autoregressive Integrated Moving Average atau VARIMA. Penelitian ini menggunakan data IHK tahun 2014-2017 sebagai data contoh untuk memprediksi nilai IHK yang akan terjadi di tahun 2018. Kelompok komoditas yang digunakan adalah pada padi-padian dan umbi-umbian dengan nilai MAPE 0,7359% serta bumbu-bumbuan dengan nilai MAPE 10,6736% (Ayudhiah, Bahri and Fitriyani, 2020). Penelitian oleh Dimashanti dan Sugiman selanjutnya menggunakan metode pengembangan dari ARIMA yaitu Seasonal ARIMA atau SARIMA. Pada penelitian ini ditemukan formula model SARIMA (1,1,1)(2,1,0) dengan nilai MSE 0,3639. Formula inilah yang digunakan untuk memprediksi nilai IHK pada Januari 2019 hingga Desember 2021. Data yang digunakan sebagai data

contoh berasal dari data IHK Januari 2014 hingga Desember 2018 (Dimashanti and Sugiman, 2021). Metode kembar *time series* memiliki performa yang baik dalam prediksi nilai IHK. Namun dalam penelitian yang dilakukan sebelumnya data yang digunakan sebagai data masukan dan data keluaran berasal dari nilai IHK itu sendiri, sehingga kompleksitas data yang digunakan cenderung kurang.

Penelitian lain melakukan prediksi IHK dengan memanfaatkan 28 komoditas makanan dan waktu sebagai data masukan. Sedangkan data keluaran yang digunakan adalah nilai IHK. Algoritma yang digunakan merupakan gabungan dari Multilayer Perceptron, LSTM, dan LSTM Bidirectional (Zahara and S. Sugianto, 2021; Zahara and Sugianto, 2021). Penelitian pertama dilakukan menggunakan Multilayer Perceptron dengan nilai RMSE 3,380. Pada tahun yang sama Zahara kembali melakukan penelitian untuk prediksi nilai IHK menggunakan gabungan dari beberapa metode untuk memperluas area eksplorasi. Hasil penelitian pada percobaan kedua menunjukkan bahwa LSTM mendapatkan RMSE terkecil yaitu 4,334 disusul dengan LSTM Bidirectional yaitu 5,630, kemudian Multilayer Perceptron yaitu 6,304.

Penelitian selanjutnya memanfaatkan Support Vector Regression (SVR) (Rohmah *et al.*, 2019). Hasil dari penelitian Rohmah memiliki tingkat kesalahan yang rendah dengan nilai MAPE kurang dari 3. Pada penelitian Rohmah dilakukan prediksi dengan memanfaatkan 34 komoditas makanan sebagai data masukan dan data IHK sebagai data keluaran. *Sample* kota yang digunakan pada penelitian Rohmah yaitu Kabupaten Banyuwangi, Kabupaten Jember, dan Kabupaten Sumenep. Nilai MAPE yang dihasilkan antara lain Kabupaten Sumenep

mempunyai nilai MAPE = 1.5508, Kabupaten Jember MAPE = 1.6823, dan Kabupaten Banyuwangi MAPE 0.851. Berdasarkan penelitian ini dapat disimpulkan bahwa SVR memiliki performa yang cukup baik dalam prediksi nilai IHK yang melibatkan data komoditas sebagai data masukannya.

Berdasarkan paparan penelitian yang dilakukan sebelumnya dapat disimpulkan bahwa mayoritas penelitian mengenai IHK masih menggunakan data IHK sebagai data masukan dan data keluaran. Meskipun hasil *error rate* terbilang rendah, namun kompleksitas data cenderung kurang beragam. Metode *Deep Learning* juga digunakan dalam proses prediksi IHK, namun tingkat kesalahan yang dihasilkan tergolong cukup tinggi. Penelitian yang dilakukan Rohmah, dkk memanfaatkan SVR dalam proses prediksi IHK dan mendapatkan hasil yang cukup memuaskan dengan rata-rata nilai MAPE kurang dari 3. Untuk menyempurnakan penelitian Rohmah, dkk maka dalam penelitian ini akan membandingkan metode SVR dengan PR. Selain itu akan dicari pula pengaruh yang terjadi antar variabel menggunakan metode Pearson Product Moments.

2.2. Keaslian Penelitian

Tabel 2. 1 Matriks Literatur Review dan Posisi Penelitian
Prediksi Indeks Harga Konsumen Komoditas Makanan di Jawa Timur menggunakan Support Vector Regression dan Polynomial Regression

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Peramalan Indeks Harga Konsumen Kota Semarang menggunakan SARIMA berbantuan Software Minitab	Dimashanti, dkk. PRISMA 4, 2021.	Mengetahui model SARIMA terbaik sehingga dapat meramalkan Indeks Harga Konsumen Kota Semarang tahun 2019-2021.	Hasil SARIMA (1,1,1) (2,1,0) dengan hasil peramalan sebanyak 36 periode ke depa dari bulan Januari 2019 sampai dengan Desember 2021.	Hanya melakukan prediksi menggunakan nilai IHK nilai masukan dan keluaran.	Pada penelitian Dimashanti, dkk hanya menggunakan data nilai IHK sebagai data latih dan data uji, sedangkan pada penelitian ini menggunakan data harga komoditas makanan sebagai data latih dan nilai IHK sebagai data uji.
2	Peramalan Indeks Harga Konsumen Indonesia menggunakan Autogressive Integrated Moving Average	Mukron, dkk. Jurnal Statistika Industri dan Komputasi, 2021.	Mengetahui nilai prediksi IHK Indonesia pada bulan selama 5 periode ke depan menggunakan metode ARIMA.	Hasil ARIMA (2,1,3) dengan nilai Mean Square sebesar 0,1744.	Hanya melakukan prediksi menggunakan nilai IHK nilai masukan dan keluaran.	Pada penelitian Mukron, dkk hanya menggunakan data nilai IHK sebagai data latih dan data uji, sedangkan pada penelitian ini menggunakan data harga komoditas makanan sebagai data latih dan nilai IHK sebagai data uji.

Tabel 2. 1 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	Peramalan Indeks Harga Konsumen Kota Mataram menggunakan Vector Autoregressive Integrated Moving Average (VARIMA)	Ayudhiah dkk. Eigen Mathematics Journal, 2020.	Meramalkan data Indeks Harga Konsumen (IHK) sub-kelompok padi-padian, umbi-umbian, dan hasilnya, serta sub-kelompok bumbu-bumbuan di Kota Mataram.	Hasil penelitian menunjukkan bahwa model terbaik adalah model VARIMA (1,1,0) dengan akurasi untuk IHK padi-padian, umbi-umbian, dan hasilnya berdasarkan nilai MAPE sebesar 0,7359% (hasil peramalan dapat dikategorikan sangat baik), sedangkan akurasi model untuk IHK bumbu-bumbuan berdasarkan nilai MAPE sebesar 10,6736% (hasil peramalan dapat dikategorikan baik).	Hanya melakukan prediksi menggunakan nilai IHK nilai masukan dan keluaran.	Pada penelitian Ayudhiah, dkk hanya menggunakan data nilai IHK sebagai data latih dan data uji, sedangkan pada penelitian ini menggunakan data harga komoditas makanan sebagai data latih dan nilai IHK sebagai data uji.

Tabel 2. 1 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
4	Prediksi Indeks Harga Konsumen Komoditas Makanan berbasis Cloud Computing menggunakan Multilayer Perceptron	Zahara, dkk. JOINTECS, 2021.	Penelitian ini membangun model peramalan dengan multi variabel masukan yaitu 28 jenis harga komoditas makanan harian sebagai nilai masukan untuk meramal nilai Indeks Harga Konsumen di kota Surabaya periode 2014 sampai 2018.	Akurasi terbaik dengan nilai RMSE 3,380 dihasilkan oleh konfigurasi 2 hidden layer, hidden layer pertama dan kedua mempunyai neuron masing-masing berjumlah 10 dengan epoch sebesar 1000.	Periode data yang digunakan hanya terdiri dari tahun 2014-2018.	Periode data yang digunakan pada penelitian Zahara, dkk adalah dari tahun 2014-2018. Sedangkan pada penelitian ini periode data yang digunakan terdiri dari tahun 2014-2021.
5	Peramalan Data Indeks Harga Konsumen berbasis Time Series Multivariate menggunakan Deep Learning	Zahara, dkk. JURNAL RESTI, 2021.	Membangun model peramalan IHK dengan algoritma deep learning yaitu LSTM, LSTM Bidirectional, dan Multilayer Perceptron dengan variasi arsitektur jumlah neuron dan epoch.	Nilai akurasi terbaik dengan nilai RMSE 3,519 didapatkan dari algoritma Bidirectional LSTM dengan konfigurasi 10 neuron dan 2000 epoch.	Periode data yang digunakan hanya terdiri dari tahun 2014-2018.	Periode data yang digunakan pada penelitian Zahara, dkk adalah dari tahun 2014-2018. Periode data yang digunakan terdiri dari tahun 2014-2021.

Tabel 2. 1 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
6	Penerapan Metode Exponential Smoothing dalam Peramalan Indeks Harga Konsumen (IHK) Provinsi Sumatera Utara	Harahap, dkk. <i>Gamma-Pi: Jurnal Matematika dan Terapan</i> , 2020.	Untuk mengetahui penerapan metode exponential smoothing dalam peramalan Indeks Harga Konsumen (IHK) Provinsi Sumatera Utara tahun 2015 – 2020.	Lebih tepat menggunakan model Additive karena memiliki error yang lebih kecil dibandingkan model Multiplicative berdasarkan nilai MAPE yaitu sebesar 0,57%.	Hanya melakukan prediksi menggunakan nilai IHK nilai masukan dan keluaran.	Pada penelitian Harahap, dkk hanya menggunakan data nilai IHK sebagai data latih dan data uji, sedangkan pada penelitian ini menggunakan data harga komoditas makanan sebagai data latih dan nilai IHK sebagai data uji.
7	Peramalan Data Indeks Harga Konsumen Kota Purwokerto menggunakan Model Fungsi Transfer Multi Masukan	Ananda, dkk. <i>Jurnal Gaussian</i> , 2020.	Menganalisis model fungsi transfer sehingga model yang terbaik adalah diproduksi untuk memprediksi IHK di Purwokerto.	Model terbaik yaitu model fungsi transfer multi masukan (2,0,0) (0,1,0) dan ARIMA (31, 0,0). Model memiliki nilai Akaike's Information Criterion (AIC) sebesar 72.42021 dan nilai MAPE sebesar 2,351591%	Hanya melakukan prediksi menggunakan nilai IHK nilai masukan dan keluaran.	Pada penelitian Ananda, dkk hanya menggunakan data nilai IHK sebagai data latih dan data uji, sedangkan pada penelitian ini menggunakan data harga komoditas makanan sebagai data latih dan nilai IHK sebagai data uji.

Tabel 2. 1 (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
8	Prediksi Indeks Harga Konsumen menggunakan Algoritma Backpropagation dengan Optimasi Particle Swarm Optimization	Muta'ali, Tesis Universitas Pembangunan Nasional "Veteran" Yogyakarta, 2021. http://eprints.upnyk.ac.id/24506/	Memprediksi IHK menggunakan Algoritma Backpropagation dengan Optimasi Particle Swarm Optimization.	Akurasi paling optimal diperoleh algoritma backpropagation dengan nilai RMSE 0.0448 dan MAPE sebesar 7.91%.	Hanya melakukan prediksi menggunakan nilai IHK nilai masukan dan keluaran.	Pada penelitian Muta'ali, dkk hanya menggunakan data nilai IHK sebagai data latih dan data uji, sedangkan pada penelitian ini menggunakan data harga komoditas makanan sebagai data latih dan nilai IHK sebagai data uji.
9	Meramal Indeks Harga Konsumen Kabupaten Jawa Timur dengan Metode Support Vector Regression Data Mining	Rohmah, dkk. SNATI, 2019.	Meramal IHK untuk bahan makanan di Kabupaten Banyuwangi, Jember, dan Sumenep.	Kabupaten Sumenep mempunyai nilai terkecil (MSE = 2.5182 dan MAPE = 1.5508) dibandingkan dengan kabupaten lainnya.	Data yang digunakan hanya berasal dari Kabupaten Banyuwangi, Jember, dan Sumenep.	Pada penelitian Rohmah, dkk, data kota yang dijadikan sebagai sampel adalah 3 kota, sedangkan pada penelitian ini data yang digunakan berasal dari Jember, Banyuwangi, Sumenep, Kediri, Malang, Probolinggo, Madiun, dan Surabaya.
10	Peramalan Indeks Harga Konsumen Kota Semarang menggunakan SARIMA berbantuan Software Minitab	Dimashanti, dkk. PRISMA 4, 2021.	Mengetahui model SARIMA terbaik sehingga dapat meramalkan Indeks Harga Konsumen Kota Semarang tahun 2019-2021.	Hasil SARIMA (1,1,1) (2,1,0) dengan hasil peramalan sebanyak 36 periode ke depan dari bulan Januari 2019 sampai dengan Desember 2021.	Hanya melakukan prediksi menggunakan nilai IHK nilai masukan dan keluaran.	Pada penelitian Dimashanti, dkk hanya menggunakan data nilai IHK sebagai data latih dan data uji, sedangkan pada penelitian ini menggunakan data harga komoditas makanan sebagai data latih dan nilai IHK sebagai data uji.

2.3. Landasan Teori

Landasan teori memuat rangkuman mengenai teori yang digunakan dalam penelitian tesis ini. Teori yang dibahas diantaranya: Indeks Harga Konsumen (IHK), Komoditas Makanan, Support Vector Regression, Polynomial Regression, dan Product Pearson Moments.

2.3.1. Indeks Harga Konsumen

Gejolak peristiwa ekonomi, masyarakat, serta perubahan harga yang terjadi beberapa waktu terakhir mengakibatkan adanya perubahan pola konsumsi masyarakat. Data pola konsumsi dibutuhkan sebagai bahan dasar penyusunan Indeks Harga Konsumen (IHK). IHK adalah suatu indeks yang menggambarkan perkembangan harga barang maupun jasa yang terjadi setelah tahun dasar (Timur, 2020). Indeks inilah yang menjadi salah satu indikator yang memberikan informasi mengenai perkembangan harga barang/jasa yang harus dibayar oleh konsumen. Adanya perhitungan IHK bertujuan untuk menemukan perubahan harga dari barang/jasa yang digunakan oleh masyarakat. Naiknya nilai IHK menggambarkan adanya inflasi yang berakibat pada kenaikan harga barang/jasa, begitu pula sebaliknya. Jika terdapat inflasi atau kenaikan harga, maka akan berdampak pula pada daya beli masyarakat. Semakin tinggi nilai inflasi, semakin rendah pula daya beli masyarakat akibat kurangnya kemampuan masyarakat dalam memenuhi kebutuhan hidupnya.

Pada tahun 2018, BPS melaksanakan Survey Biaya Hidup (SBH) 2018 di 90 kota di Indonesia, yang terdiri dari 34 ibukota dan 56 kabupaten/kota dengan tingkat perekonomian yang relatif tinggi. Berdasarkan SBH tersebut, Jawa Timur

memiliki 8 kabupaten/kota yang termasuk di dalamnya, antara lain: Kabupaten Jember, Kabupaten Banyuwangi, Kabupaten Sumenep, Kota Kediri, Kota Malang, Kota Probolinggo, Kota Madiun, dan Kota Surabaya (Timur, 2020). Hasil dari SBH 2018 menetapkan harga konsumen pada seluruh komoditas yang dikonsumsi oleh masyarakat di Indonesia.

Dalam penentuan nilai IHK, tentu tidak terlepas dari Harga Konsumen. Harga konsumen adalah harga transaksi yang terjadi antara penjual yang merupakan pedagang eceran dan pembeli, yang dilakukan dengan cara mengecor dan dengan pembayaran tunai. Yang dimaksud eceran disini adalah pembelian yang dilakukan dengan satuan terkecil untuk dipakai. Contohnya untuk sayuran dengan satuan ikat, beras dengan satuan kg/liter, emas dengan satuan gram, dan sebagainya. Pada implementasinya terkadang barang ada yang dijual secara kemasan, ditambah Pajak Pertambahan Nilai (PPN), dan sebagainya. Namun harga konsumen yang dicantumkan disini benar-benar merupakan harga asal, tanpa memperhatikan kemasan maupun PPN.

Setelah memiliki data harga konsumen, selanjutnya IHK akan diolah menggunakan formula Laspeyres. Formula Laspeyres dapat dilihat pada Persamaan 2.1.

$$I_n = \frac{\sum(P_n \times Q_0)}{\sum(P_0 \times Q_0)} \times 100 \dots\dots\dots (2.1)$$

Dimana:

I_n = Indeks bulan ke n

P_n = Harga tahun berjalan

P_0	=	Harga tahun dasar
Q_0	=	Kuantitas barang pada tahun sebelumnya
$P_{ot} \cdot Q_{ot}$	=	Nilai konsumsi jenis barang I pada bulan dasar
K	=	Banyaknya jenis barang paket komoditas dalam sub kelompok

2.3.2. Komoditas Makanan

Dari 8 kota/kabupaten di Jawa Timur yang termasuk dalam SBH 2018, terdapat total 563 komoditas. Pembagiannya antara lain Kabupaten Jember sebanyak 358, Kabupaten Banyuwangi sebanyak 315 komoditas, Kabupaten Sumenep sebanyak 320 komoditas, Kota Kediri sebanyak 373 komoditas, Kota Malang sebanyak 382 komoditas, Kota Probolinggo sebanyak 340 komoditas, Kota Madiun sebanyak 335 komoditas, dan Kota Surabaya sebanyak 335 komoditas.

Berdasarkan SBH 2018, jumlah kelompok komoditas berubah menjadi 11 kelompok dengan 43 sub kelompok. Dalam penelitian ini akan digunakan data dari kelompok Makanan, Minuman, dan Tembakau. Kelompok Makanan, Minuman, dan Tembakau memiliki 4 sub kelompok. Fokus yang digunakan pada penelitian ini adalah pada sub kelompok makanan atau komoditas makanan.

Dari keseluruhan komoditas yang telah disebutkan, diambil 34 komoditas yang mayoritas digunakan oleh masyarakat di Jawa Timur. Berdasarkan penelitian dari Rohmah, dkk komoditas tersebut diantaranya Beras Bengawan, Beras Mentik, Beras IR64, Gula Pasir Dalam Negeri, Minyak Goreng Bimoli Botol/Kemasan 2 ltr, Minyak Curah, Daging Sapi Murni, Daging Ayam Broiler,

Daging Ayam Kampung, Telur Ayam Ras/Petelur, Telur Ayam Kampung, Susu Kental Manis Merk Bendera, Susu Kental Manis Merk Indomilk, Susu Bubuk Merk Bendera (Instant), Susu Bubuk Merk Indomilk (Instant), Jagung Pipilan Kering, Garam yodium jenis Bata, Garam yodium jenis Halus, Tepung Terigu Segitiga Biru (Kw Medium), Kacang Kedelai Eks Impor, Mie Instant Indomie Rasa Kari Ayam, Cabe Biasa, Cabe Rawit, Bawang Merah, Bawang Putih, Ikan Asin Teri, Kacang Hijau, Kacang Tanah, Ketela Pohon, Kol/Kubis, Kentang, Tomat, Wortel, dan Buncis (Rohmah *et al.*, 2019).

2.3.3. Support Vector Regression

Algoritma Support Vector (SV) pertama kali dikembangkan pada tahun 60-an oleh Vapnik dan Learner (1963) juga Vapnik dan Chervonenkis (1964). Dalam prosesnya algoritma ini telah dikembangkan selama 30 tahun. Algoritma SV secara singkat merupakan suatu algoritma mesin pembelajaran yang digunakan untuk melakukan generalisasi data-data yang tidak terlihat. Saat ini, algoritma SV lebih cenderung dikembangkan oleh AT&T Bell Laboratories, dan memiliki orientasi yang lebih baik dalam terhadap aplikasi dunia nyata (Sande and Privalsky, 1996).

Algoritma SV memiliki keandalan yang baik dalam proses klasifikasi bersama Support Vector Machine (SVM). Meskipun belum seterkenal pendahulunya yaitu SVM, untuk kasus regresi, Support Vector Regression (SVR) juga merupakan algoritma yang handal karena dapat mengatasi *overfitting data* dengan baik. *Overfitting* merupakan sebuah kondisi ketika data berada pada proses

pelatihan dan menghasilkan akurasi prediksi yang hampir sempurna (Maulana, Setiawan and Dewi, 2019).

2.3.3.1. Ide Dasar

Sama seperti konsep yang dimiliki oleh SVM, algoritma SVR memiliki tujuan untuk mencari *hyperplane* terbaik berupa fungsi regresi yang memiliki nilai kesalahan terkecil dengan batas (margin) maksimal. Keluaran yang dihasilkan berupa bilangan riil maupun bilangan kontinu (Asyiva, 2019).

Misalnya terdapat data pelatihan $\{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\}$ dengan $x_i \in \mathbb{R}^d$ adalah vector input dari data ke- i , dimana $i = 1, 2, 3, \dots, n$. Sementara d adalah dimensi dan y_i adalah nilai target. Persamaan umum dalam SVR ditunjukkan dalam Persamaan 2.1.

$$f(x) = \langle w, x \rangle + b \quad (2.1)$$

Berdasarkan persamaan di atas, $\langle \cdot \rangle$ merupakan hasil kali yang berada di X . Fungsi dari perkalian tersebut adalah agar didapatkan generalisasi yang baik untuk fungsi regresi $f(x)$ dengan cara meminimalkan w menggunakan penyelesaian masalah optimasi berdasarkan Persamaan 2.2.

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

Dengan syarat:

$$y_i - \langle w, x \rangle - b \leq \epsilon$$

$$\langle w, x \rangle + b - y_i \leq \epsilon$$

Pada Persamaan 3 memiliki asumsi seluruh titik berada dalam rentang $f(x) \pm \varepsilon$. Jika terdapat titik yang keluar dari rentang tersebut, maka diperlukan kondisi khusus yang dinamakan *soft margin* atau *variable slack* dalam $(\xi_i + \xi_i^*)$, sehingga persamaannya berubah menjadi seperti pada Persamaan 2.3.

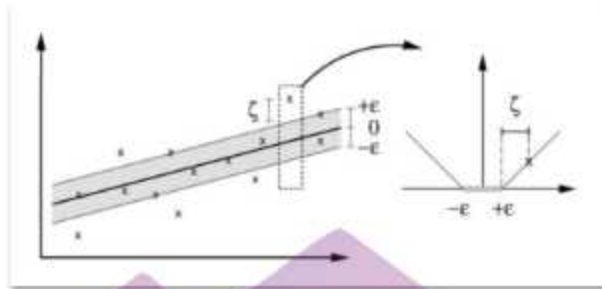
$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \dots\dots\dots (2.3)$$

Dengan syarat:

$$\begin{aligned} y_i - \langle w, x \rangle - b &\leq \varepsilon + \xi_i \\ \langle w, x \rangle + b - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned}$$

Nilai $C > 0$ merupakan nilai pinalti yang diakibatkan adanya data yang keluar dari batas toleransi terhadap fungsi $f(x)$. Nilai ini juga menginisiasi tingkat deviasi dari batas ε yang dapat ditoleransi. Jadi seluruh nilai yang masuk dan lebih dari ε akan dikenakan pinalti sebesar C . Formula tersebut biasa disebut dengan *ε -insensitive loss function* yang ditunjukkan pada Persamaan 2.4:

$$|\xi_i|_C = \begin{cases} 0 & \text{if } |\xi_i| \leq \varepsilon \\ |\xi_i| - \varepsilon & \text{if } |\xi_i| > \varepsilon \end{cases} \dots\dots\dots (2.4)$$



Gambar 2. 1 Pengaturan Soft Margin Loss untuk Linear SVM

Sumber: (Bernhard Scholkopf, 2018)

Nilai *variable slack* yang tinggi dapat mengakibatkan kesalahan dalam perhitungan. Jika nilai ϵ kecil dan nilai *variable slack* tinggi, maka nilai akurasi akan menjadi tinggi. Sedangkan jika nilai ϵ tinggi dan nilai *variable slack* rendah, maka nilai akurasi akan menjadi rendah.

Penentuan nilai parameter w dan b dapat dilakukan menggunakan pemrograman kuadrat. Pemrograman kuadrat dilakukan dengan cara meminimumkan fungsi kuadrat dalam pertidaksamaan linear. Masalah ini dapat diatasi dengan *Lagrange Multiplier* pada Persamaan 2.5.

$$L = \left(\frac{1}{2}\|w\|^2 + c \sum_{i=1}^l (\xi_i + \xi_i^*) - \left(\sum_{i=1}^l \alpha_i (\epsilon + \xi_i - y_i + \langle w, x \rangle + b)\right) - \sum_{i=1}^l \alpha_i^* (\epsilon + \xi_i + y_i - \langle w, x \rangle - b)\right) - \left(\sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*)\right) \dots \dots \dots (2.5)$$

Untuk mendapatkan solusi optimal, diberikan turunan parsial L untuk w , b , ξ_i , ξ_i^* . Turunan tersebut ditunjukkan pada Persamaan 2.6, 2.7, 2.8, dan 2.9.

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i = 0 \dots \dots \dots (2.6)$$

$$\frac{\partial L}{\partial b} = w - \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \dots \dots \dots (2.7)$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0 \quad (2.8)$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0 \quad (2.9)$$

Persamaan 2.6, 2.7, 2.8, dan 2.9 kemudian disubstitusikan ke dalam Persamaan 2.4 sehingga diperoleh persamaan seperti dalam Persamaan 2.10.

$$-\frac{1}{2} \sum_{i=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - (\alpha_j^*)) (x_i \cdot x_j) - \varepsilon \sum_{i=1}^l (\alpha_i - \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \quad (2.10)$$

Dengan syarat:

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0,$$

$$0 \leq \alpha_i \leq C \quad \forall i = 1, 2, \dots, N$$

$$0 \leq \alpha_i^* \leq C \quad \forall i = 1, 2, \dots, N$$

Berdasarkan turunan rumus pada Persamaan 2.10, didapatkan variabel utama yaitu α_i dan α_i^* . Solusi dari masalah ini diturunkan dari vektor w kemudian dimasukkan dalam substitusi fungsi $f(x)$.

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \cdot x_i \quad (2.11)$$

Setelah itu dilanjutkan dengan memaksimalkan nilai b berdasarkan kondisi:

$$\alpha_i (\varepsilon + \xi_i + y_i - \langle w, x \rangle - b) = 0 \quad (2.12)$$

$$\alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w, x \rangle - b) = 0$$

$$(C - \alpha_i) \xi_i = 0 \quad (2.13)$$

$$(C - \alpha_i^*) \xi_i^* = 0$$

Sehingga didapatkan persamaan sebagai berikut:

$$b = y_i - \sum_{i=1}^l (\alpha_i - \alpha_i^*) \cdot (x_i \cdot x_j) + \varepsilon \quad \text{dan} \quad b^* = y_i - \sum_{i=1}^l (\alpha_i - \alpha_i^*) \cdot (x_i \cdot x_j) - \varepsilon$$

Berdasarkan Persamaan 2.13, didapatkan $C = \alpha_i$, maka berakibat pada nilai w yang dihitung bergantung bergantung pada nilai *support vector*. Setelah ditemukan nilai w dan b , maka dilakukan substitusi untuk Persamaan 2.10 dan Persamaan 2.11. Substitusi ini akan menghasilkan persamaan baru seperti pada Persamaan 2.14.

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \cdot \langle x_i - x_j \rangle + b \dots\dots\dots (2.14)$$

Persamaan 2.13 berguna untuk menghitung persamaan linear. Jika terdapat permasalahan yang non-linear maka nilai x_i dan x_j ditransformasikan dulu dalam *feature space* dengan cara memetakan vektor x_i dan x_j dalam fungsi $\Phi : \mathbb{R} \rightarrow \mathbb{R}^Z$, sehingga menghasilkan persamaan:

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \Phi(x_i) \dots\dots\dots (2.15)$$

Sehingga persamaan akhir didapatkan:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \Phi(x_i) \Phi(x_j) + b \dots\dots\dots (2.16)$$

Jika vektor x_i dan x_j telah ditransformasi menggunakan fungsi Φ dan berada dalam *feature space*, maka fungsi transformasinya dapat ditampilkan dengan fungsi kernel K seperti pada Persamaan 2.17:

$$K(x_i, x_j) = \Phi(x_i) \Phi(x_j)$$

Sehingga didapatkan persamaan akhir seperti pada Persamaan 2.17. Fungsi $K(x_i, x_j)$, merupakan fungsi yang kerap kali digunakan dalam *Support Vector*, baik SVR maupun SVM.

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \dots \dots \dots (2.17)$$

2.3.3.2. Fungsi Kernel

SVR juga mendukung permasalahan data non-linier dengan bantuan dari Fungsi Kernel. Fungsi kernel pada umumnya digunakan untuk data yang memiliki dimensi tinggi. Hal ini dilakukan dengan cara mengganti *inner product*, yaitu x_i dan x_j dengan fungsi kernel (Maulana, Setiawan and Dewi, 2019).

Terdapat tiga jenis kernel yang dikenal dalam Support Vector, yaitu Kernel Linear, Polynomial, dan Radial Basic Function (RBF). Dalam penelitian ini akan digunakan Kernel BF. Kernel ini digunakan karena handal dalam mengatasi data yang berdimensi tinggi, serta terbukti unggul dari kernel-kernel yang lainnya (Asyiva, 2019). Formula dalam RBF dapat dilihat pada Persamaan 2.18:

$$K(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2}\right)(x_i - x_j) \dots \dots \dots (2.18)$$

2.3.3.3. Grid Search Optimization

Untuk mendapatkan hasil uji coba yang akurat, diperlukan pula *hyperparameter* yang tepat. Hyperparameter merupakan suatu parameter yang mengatur jalannya pembelajaran mesin. *Hyperparameter* yang dapat digunakan pada model SVR adalah Grid Search Optimization (GSO).

Metode ini bekerja dengan cara membagi jangkauan parameter yang ada ke dalam *grid* yang melintasi semua titik demi mencapai parameter optimal (Asyiva, 2019). Tujuan dari GSO adalah untuk menentukan kombinasi yang menghasilkan performa model terbaik untuk dijadikan model prediksi.

2.3.4. Polynomial Regression

Polynomial Regression (PR) merupakan suatu jenis regresi yang bekerja pada hubungan lengkung (kurva) antara variabel bebas dan variabel terikat (Eka *et al.*, 2021). Cara kerja algoritma regresi ini adalah dengan cara menjumlahkan pengaruh pada variabel x yang dipangkatkan sampai ke orde p (Susianto, 2016). PR merupakan kelanjutan dari algoritma regresi linear berganda. Perbedaannya adalah bahwa pada PR diperlukan derajat PRnya.

Namun perlu diketahui pada saat pemilihan derajat PR, salah satu masalah yang sering terjadi adalah sulitnya menentukan derajat yang tepat. PR memiliki derajat hingga 6 derajat atau yang biasa disebut orde. Jika derajat yang dipilih rendah, maka galatnya akan besar. Jika derajat yang dipilih tinggi, maka galat akan menjadi kecil atau bahkan tidak memiliki galat.

Tujuan utama yang ingin dicapai dalam PR adalah membangun model yang sederhana untuk melakukan proses prediksi. Semakin tinggi derajat polinomial yang digunakan, maka semakin kompleks pula model yang didapat. Namun model yang kompleks juga tidak mencerminkan keandalan suatu model prediksi. Karena nilai galat yang tercipta mungkin hanya baik pada data yang sedang digunakan, dan

memungkinkan bahwa tidak akan sama hasilnya jika digunakan untuk data yang lain. Secara umum, PR ditulis pada Persamaan 2.19 dan Persamaan 2.20:

$$Y = b_0 + b_1X + b_2X^2 + \dots + b_nX^n + \mathcal{E} \dots\dots\dots (2.19)$$

Apabila Y merupakan nilai yang diprediksi:

$$Y = b_0 + b_1X + b_2X^2 + \dots + b_nX^n \dots\dots\dots (2.20)$$

Dimana:

- Y : Variabel yang diprediksi
 b_0 : Intersep
 b_1, b_2, \dots, b_n : Koefisien regresi
 X : Variabel bebas
 n : Orde atau derajat polinomial
 \mathcal{E} : Galat

PR merupakan hasil modifikasi dari regresi linier berganda. Dari regresi linier berganda selanjutnya dicari turunan parsial dari SSE terhadap koefisien beta dan menjadikannya sama dengan nol, sehingga didapatkan formula pada Persamaan 2.21:

$$X(p) = \begin{bmatrix} n & \sum x_i & \dots & \sum x_i^p \\ \sum x_i & \sum x_i^p & \dots & \sum x_i^{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_i^p & \sum x_i^{p+1} & \dots & \sum x_i^{p+2} \end{bmatrix} \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} Y(p) = \begin{bmatrix} \sum y_i \\ \sum y_i x_i \\ \vdots \\ \sum y_i x_i^p \end{bmatrix} \dots\dots\dots (2.21)$$

Atau jika disederhanakan menjadi pada Persamaan 2.22 dan Persamaan 2.23:

$$X(p)\beta = Y(p) \dots\dots\dots (2.22)$$

$$\beta = X(p)^{-1}Y(p) \dots\dots\dots (2.23)$$

Dimana:

X = Variabel bebas

p = Derajat

n = Jumlah data

y = Variabel terikat ke I

β = Variabel regresi

2.3.5. Product Pearson Moments

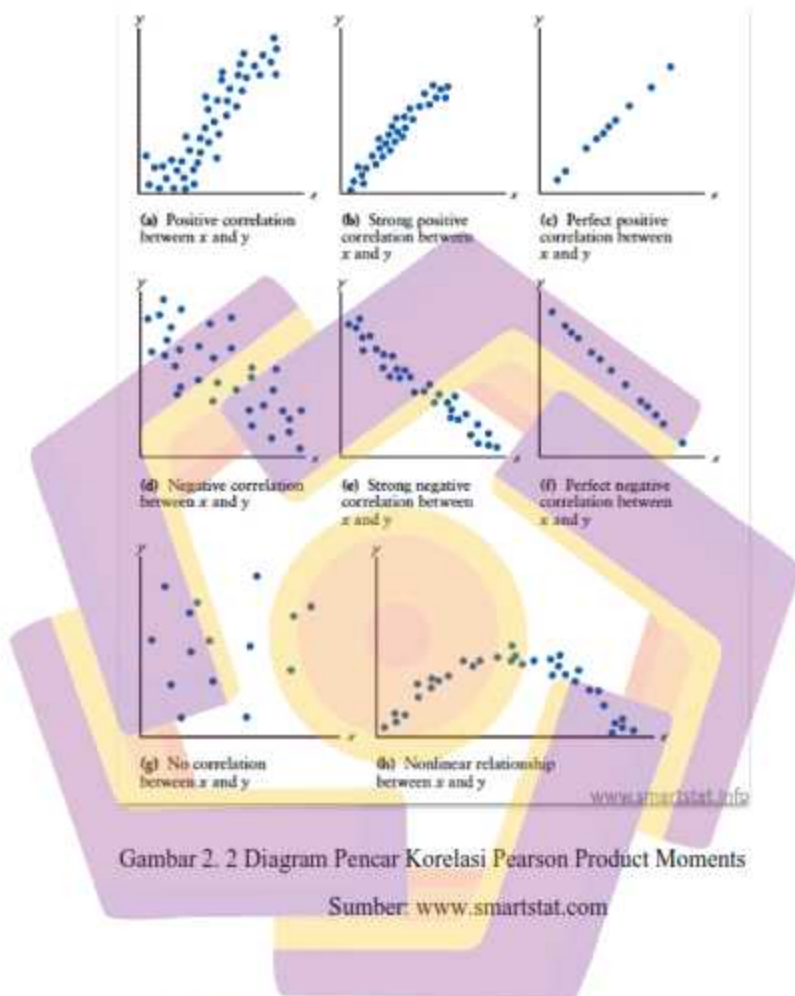
Untuk mengetahui hubungan yang terbentuk antar variabel, digunakan suatu metode yang disebut analisa korelasi. Analisa korelasi dibagi menjadi dua jenis berdasarkan data yang digunakan, yaitu analisis korelasi untuk data berskala interval dan analisis data untuk data berskala ordinal (Erfiana, 2015).

Pada penelitian ini akan dilakukan analisis korelasi untuk data berskala interval menggunakan metode Pearson Product Moments. Korelasi ini digunakan untuk mengukur keeratan hubungan di antara populasi yang ada untuk mengetahui arah hubungan yang terjadi. Analisa korelasi menggunakan Product Pearson Moments ditunjukkan pada Persamaan 2.24:

$$r = \frac{n \sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{\sqrt{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \sqrt{n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2}} \dots\dots\dots (2.24)$$

Nilai korelasi yang terbentuk berkisar antara 1 hingga -1. Jika nilai korelasi (r) semakin mendekati 1 atau -1, maka berarti hubungan yang terbentuk antar variabel semakin kuat. Sebaliknya jika r semakin mendekati nilai 0, maka hubungan yang terbentuk antar variabel semakin lemah. Nilai r positif menunjukkan hubungan yang searah (jika X naik, maka Y naik), sedangkan nilai r negative menunjukkan hubungan terbalik (jika X naik, maka Y turun) (Gambar 2. 2). Pedoman interpretasi koefisien korelasi menurut (Sugiyono, 2012), sebagai berikut:

1. $0,00 - 0,199 =$ Korelasi sangat rendah
 2. $0,20 - 0,399 =$ Korelasi rendah
 3. $0,40 - 0,599 =$ Korelasi sedang
 4. $0,60 - 0,799 =$ Korelasi kuat
 5. $0,80 - 1,000 =$ Korelasi sangat kuat
- 



Gambar 2. 2 Diagram Pencar Korelasi Pearson Product Moments

Sumber: www.smartstat.com

2.3.6. Evaluasi Hasil Prediksi

Beberapa metode yang bisa digunakan untuk mengevaluasi model prediksi yang telah terbentuk. Evaluasi dilakukan untuk mencari tahu seberapa tingkat kesalahan yang terbentuk dari model prediksi. Semakin kecil tingkat kesalahan yang terbentuk maka semakin baik modelnya. Pada penelitian ini akan digunakan

3 metode evaluasi, yakni Mean Square Error (MSE), Root Mean Square Error (RMSE), dan Mean Absolute Percentage Error (MAPE).

2.3.6.1. Mean Square Error (MSE)

Metode evaluasi model menggunakan MSE bekerja dengan cara menghitung rata-rata selisih kuadrat antara nilai hasil prediksi dan nilai yang sebenarnya (Hasan, 2020). MSE merupakan metode evaluasi yang banyak digunakan dalam proses prediksi karena dapat memberikan nilai kesalahan yang kecil, namun tidak menutup kemungkinan juga MSE memberikan nilai kesalahan yang sangat besar bergantung pada data yang digunakan (Maricar, 2019). Formula dalam MSE ditunjukkan dalam Persamaan 2.25:

$$MSE = \frac{1}{n} \sum_{t=1}^n |X_t - F_t|^2 \quad (2.25)$$

Dimana:

MSE = Mean Square Error

X_t = Nilai aktual pada periode t

F_t = Nilai prediksi pada periode t

n = Banyaknya data

2.3.6.2. Root Mean Square Error (RMSE)

Beda dengan MSE, jika dalam RMSE proses evaluasi model prediksi dilakukan dengan cara mencari nilai rata-rata dari jumlah kuadrat kesalahan yang dihasilkan oleh suatu model prediksi (Aryani *et al.*, 2020). RMSE dikenal sebagai metode evaluasi yang digunakan untuk membandingkan metode-metode evaluasi lain yang digunakan untuk menentukan metode yang paling akurat (Wahyu

Widayati, 2013). Nilai RMSE yang rendah menunjukkan variasi nilai prediksi yang mendekati sempurna. Formula RMSE ditunjukkan dalam Persamaan 2.26:

$$RMSE = \sqrt{\frac{(\text{eyt}-f_t)^2}{n}} \dots\dots\dots (2.26)$$

Dimana:

Y_t = Nilai aktual pada periode t

F_t = Nilai prediksi pada periode t

n = Banyaknya data

2.3.6.3. Mean Absolute Percentage Error (MAPE)

Metode ini berfungsi untuk mencari rata-rata perbedaan absolut yang terbentuk dari hasil prediksi dan nilai aktual yang ditunjukkan dalam bentuk persen (Nabillah and Ranggadara, 2020). Metode ini banyak digunakan untuk melakukan evaluasi pada model prediksi karena dapat memberikan hasil evaluasi sekaligus kriteria performa evaluasinya. Dengan kriteria yang jelas tersebut dapat memudahkan peneliti untuk mengetahui seberapa baik performa suatu model melalui nilai MAPE yang terbentuk. Evaluasi menggunakan MAPE dapat menunjukkan tingkat akurasi antara nilai hasil prediksi dan nilai sebenarnya. Formula MAPE ditunjukkan pada Persamaan 2.27:

$$MAPE = \left(\frac{y_t - f_t}{f_t}\right) \times 100\% \dots\dots\dots (2.27)$$

Dimana:

Y_t = Nilai aktual pada periode t

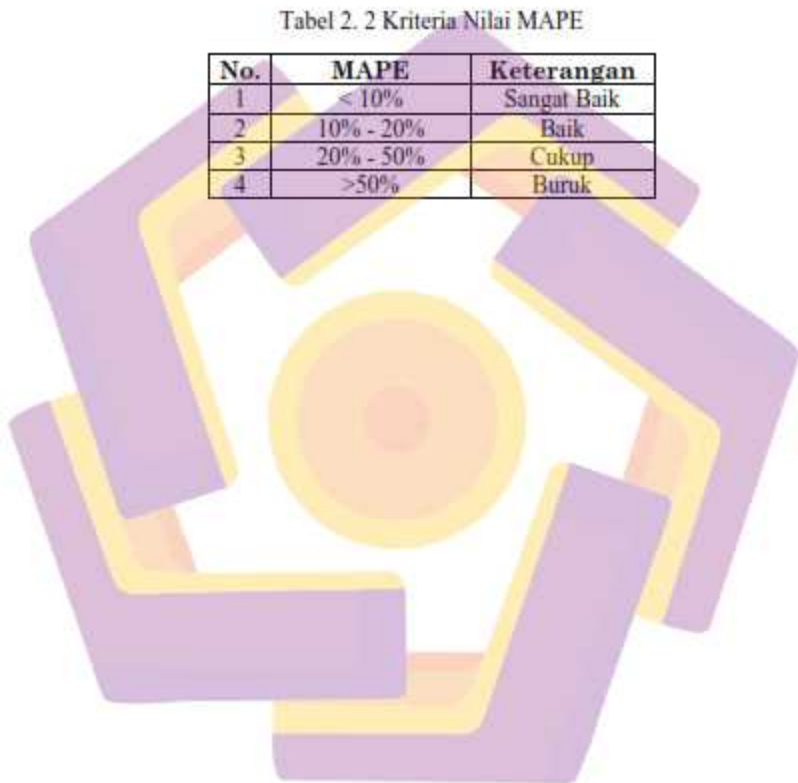
F_t = Nilai prediksi pada periode t

n = Banyaknya data

Pada implementasinya, nilai MAPE memiliki kriteria untuk proses kategorisasinya. Kriteria nilai MAPE dapat dilihat pada Tabel 2. 2 (Chang, P.-C., Wang, Y.-W. & Liu, 2007).

Tabel 2. 2 Kriteria Nilai MAPE

No.	MAPE	Keterangan
1	< 10%	Sangat Baik
2	10% - 20%	Baik
3	20% - 50%	Cukup
4	>50%	Buruk



BAB III METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Jenis penelitian yang digunakan adalah eksperimen. Eksperimen dilakukan dengan melibatkan serangkaian tindakan. Eksperimen yang dilakukan dalam penelitian ini bertujuan untuk menerapkan algoritma SVR dan PR dalam prediksi nilai IHK. Hasil penerapan algoritma kemudian akan dibandingkan untuk menemukan algoritma terbaik dengan tingkat kesalahan terkecil.

Sifat penelitian yang dilakukan adalah sifat kausal. Sifat penelitian kausal bertujuan untuk mencari hubungan sebab-akibat antara dua variabel atau lebih. Dalam penelitian ini akan dicari hubungan sebab-akibat dari harga komoditas makanan sebagai variabel independen terhadap IHK sebagai variabel terikat. Pendekatan yang digunakan dalam penelitian adalah kuantitatif. Pendekatan ini didasarkan pada data yang dipakai, yaitu data numerik. Data masukan yang berasal dari harga komoditas makanan dan nilai IHK sama-sama berbentuk numerik atau angka.

Obyek penelitian yang digunakan adalah data harga komoditas makanan di Jawa Timur dan data IHK. Proses pertama dilakukan untuk mencari korelasi antara variabel bebas dan variabel terikat menggunakan Pearson Product Moments. Dari hubungan antara variabel yang ada apakah ada keterkaitan. Setelah pencarian korelasi dilakukan dilanjutkan dengan proses prediksi. Data kemudian digunakan untuk proses prediksi menggunakan metode SVR dan PR. Hasil prediksi

menggunakan SVR dan PR kemudian di evaluasi untuk menentukan seberapa baik performanya dengan cara menentukan tingkat kesalahan. Semakin kecil tingkat kesalahannya, maka semakin baik pula performa model yang digunakan.

3.2. Metode Pengumpulan Data

Metode pengumpulan data dilakukan dengan cara observasi. Data yang digunakan merupakan jenis data sekunder yang didapat dan dikumpulkan dari harga komoditas makanan di situs <https://siskaperbapo.jatimprov.go.id/>. Data yang diambil merupakan data harian harga komoditas makanan. Periode data yang diambil berada pada rentang 1 Januari 2014 hingga 31 Desember 2020. Data ini digunakan sebagai data masukan.

Selain menggunakan data harga komoditas makanan, digunakan pula data IHK dari situs <https://jatim.bps.go.id/>. Data ini berbentuk publikasi yang diterbitkan oleh BPS setiap tahunnya. Data kemudian diolah ke dalam bentuk tabel. Data ini digunakan sebagai data keluaran.

3.3. Metode Analisis Data

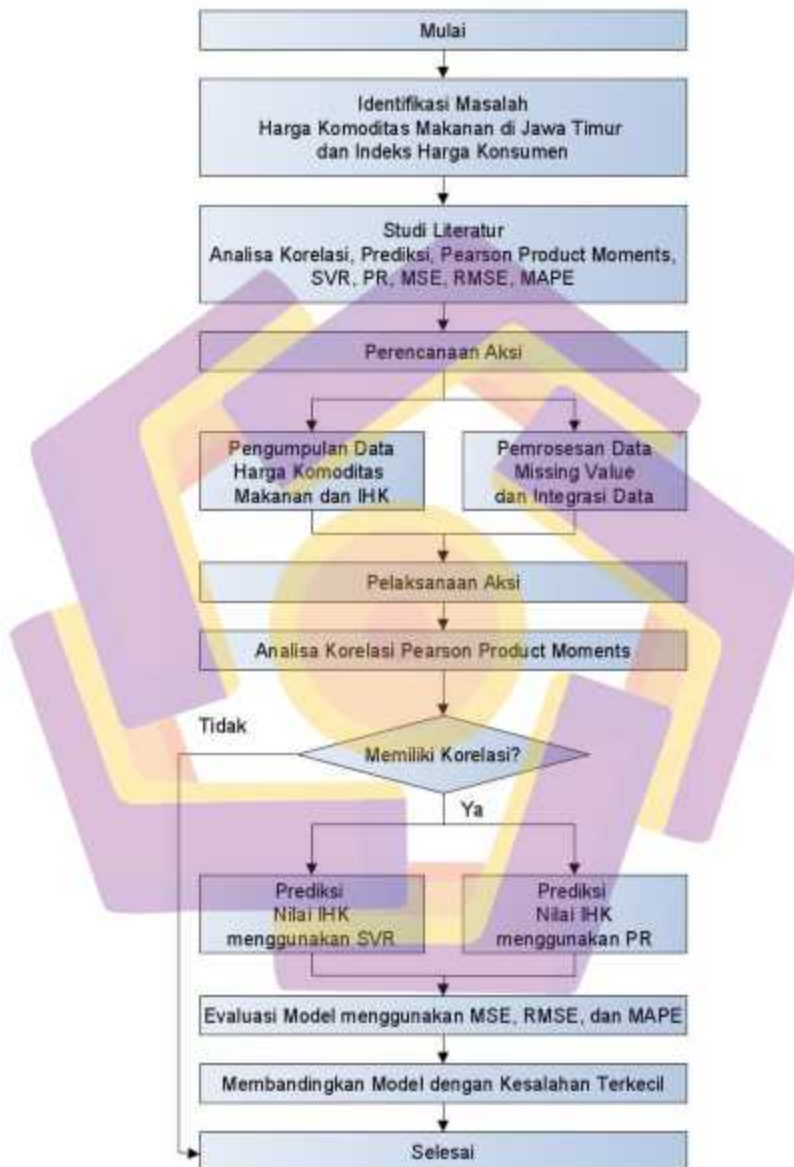
Metode analisis data akan dijelaskan melalui langkah-langkah penelitian sebagai berikut:

1. Melakukan identifikasi masalah terkait Harga Komoditas Makanan dan Nilai IHK pada 8 kota/kabupaten di Provinsi Jawa Timur;
2. Melakukan studi literatur dan konsep mengenai teori Analisa Korelasi, Prediksi, Pearson Product Moment, SVR, PR, MSE, RMSE, dan MAPE;

3. Melakukan Perencanaan Aksi. Kegiatan ini meliputi dua tugas, yaitu Pengumpulan Data dan Pemrosesan Data;
4. Pengumpulan Data dilakukan untuk mengumpulkan data harga 34 komoditas makanan harian dan juga data nilai IHK;
5. Pemrosesan data dilakukan untuk membersihkan data dari *missing value* dan melakukan integrasi data antara harga komoditas makanan dan nilai IHK-nya;
6. Dilakukan proses analisa korelasi antara variabel harga komoditas makanan dan nilai IHK. Metode Pearson Product Moments digunakan untuk melakukan analisa korelasi. Jika terdapat korelasi, maka data tersebut dapat digunakan untuk melakukan proses prediksi menggunakan metode regresi;
7. Dilanjutkan dengan melakukan prediksi menggunakan model SVR dan PR. Proses implementasi model dilakukan dengan proses evaluasi menggunakan MSE, RMSE, dan MAPE;
8. Melakukan perbandingan tingkat kesalahan dari hasil pengujian untuk mengetahui algoritma terbaik;
9. Penarikan kesimpulan.

3.4. Alur Penelitian

Alur penelitian yang digunakan mengadaptasi dari The Action Research Cycle. Metode ini diadopsi dari Baskerville (1999) dan dikutip dari (Azhar, 2007). Alur penelitian ditunjukkan pada Gambar 3.1:



Gambar 3. 1 Alur Penelitian

Penjelasan alur penelitian dijelaskan sebagai berikut:

1. Diagnosa

Pada tahap diagnosa, dilakukan proses identifikasi masalah terkait adanya masalah inflasi dan deflasi yang ada pada komoditas makanan di Jawa Timur. Salah satu aspek penting untuk mengetahuinya adalah dengan menghitung nilai IHK. Selanjutnya dilakukan proses studi literatur yang terkait dengan masalah yang sudah diidentifikasi. Literatur yang digunakan yaitu terkait proses analisa korelasi (Pearson Product Moments), prediksi (SVR dan PR), serta evaluasi (MSE, RMSE, MAPE).

2. Perencanaan Aksi

Pada tahap perencanaan aksi dilakukan proses pengumpulan data yang digunakan dalam penelitian. Data yang digunakan adalah data harga komoditas makanan yang bersumber dari Website Siskaperbapo Jawa Timur, serta data nilai IHK yang berasal dari publikasi rutin Website BPS Jawa Timur. Data yang digunakan berasal dari periode 2014-2020.

Langkah selanjutnya adalah melakukan pra proses data. Pertama akan dilakukan pembersihan data yang memiliki *value* kosong. Selanjutnya dilakukan inisialisasi dan normalisasi data agar data siap untuk diolah pada proses selanjutnya.

3. Pelaksanaan Aksi

Pada tahap ini dilakukan proses implementasi data yang sudah dikumpulkan. Pertama dicari tahu dulu korelasi antara variabel harga komoditas makanan sebagai variabel bebas dengan nilai IHK sebagai variabel terikat. Metode

yang digunakan adalah Pearson Product Moments. Setelah melakukan Analisa korelasi, dilanjutkan dengan implementasi metode SVR dan PR untuk proses prediksi. Proses implementasi dilakukan menggunakan Bahasa Pemrograman Python menggunakan Jupyter Notebook.

4. Evaluasi

Setelah tahap *action taking* dilakukan, proses selanjutnya adalah evaluasi. Proses evaluasi dilakukan untuk mencari seberapa baik performa model yang terbentuk. Model dengan tingkat kesalahan yang paling rendah akan memiliki performa paling baik. Metode yang digunakan adalah MSE, RMSE, dan MAPE.

5. Menentukan Pembelajaran

Tahap terakhir dilakukan untuk mengambil kesimpulan atas penelitian yang dilakukan. Pada proses ini akan ditunjukkan metode dengan performa paling baik yang dapat direkomendasikan sebagai metode terbaik untuk prediksi nilai IHK.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Perencanaan Aksi

Pada sub bab ini akan dibahas mengenai perencanaan aksi untuk penyelesaian masalah penelitian. Perencanaan aksi akan terdiri dari beberapa proses, diantaranya: pengumpulan data dan pra proses data.

4.1.1. Pengumpulan Data

Pengumpulan data dibagi menjadi dua sumber, yaitu Website Siskaperbapo Jawa Timur (<https://siskaperbapo.jatimprov.go.id/>) untuk data harga komoditas makanan. Sedangkan data Indeks Harga Konsumen (IHK) bersumber dari Website BPS Provinsi Jawa Timur (<https://jatim.bps.go.id/>).

a. Data Harga Komoditas Makanan

Data tersebut terdiri dari 34 komoditas, yaitu Beras Bengawan, Beras Mentik, Beras IR64, Gula Pasir Dalam Negeri, Minyak Goreng Bimoli Botol/Kemasan 2 ltr, Minyak Curah, Daging Sapi Murni, Daging Ayam Broiler, Daging Ayam Kampung, Telur Ayam Ras/Petelur, Telur Ayam Kampung, Susu Kental Manis Merk Bendera, Susu Kental Manis Merk Indomilk, Susu Bubuk Merk Bendera (Instant), Susu Bubuk Merk Indomilk (Instant), Jagung Pipilan Kering, Garam yodium jenis Bata, Garam yodium jenis Halus, Tepung Terigu Segitiga Biru (Kw Medium), Kacang Kedelai Eks Impor, Mie Instant Indomie Rasa Kari Ayam, Cabe Biasa, Cabe Rawit, Bawang Merah, Bawang Putih, Ikan Asin Teri, Kacang Hijau, Kacang Tanah, Ketela Pohon, Kol/Kubis, Kentang, Tomat, Wortel, dan

Buncis. Data harga komoditas makanan terdiri dari 8 kabupaten/kota di Jawa Timur. Periode data yang digunakan yaitu data harian harga komoditas makanan dari tahun 2014-2020. Total data yang dikumpulkan yaitu 20.453 data. Contoh data dapat dilihat dalam Tabel 4. 1.

Tabel 4. 1 Contoh Data Harian Komoditas Makanan

DAERAH	KOMODITAS	01/01/2014	01/02/2014	01/03/2014	...	31/12/2020
Jember	Bawang Merah	33.000	33.000	33.000	...	33.333
Jember	Bawang Putih	11.000	11.000	10.000	...	20.000
Jember	Beras Bengawan	8.500	8.500	8.500	...	11.000
Jember	Beras IR64	7.500	7.500	7.500	...	10.167
Jember	Beras Mentik	9.600	9.600	9.600	...	12.167
Jember	Buncis	7.000	7.000	8.000	...	10.667
Jember	Cabe Besar Biasa	33.000	33.000	33.000	...	36.333
Jember	Cabe Rawit	23.000	23.000	23.000	...	46.667
Jember	Daging Ayam Broiler	24.000	24.000	25.000	...	33.000
Jember	Daging Ayam Kampung	35.000	35.000	35.000	...	53.333
Jember	Daging Sapi Mum	90.000	90.000	90.000	...	110.000
Jember	Garam yodium jenis Bata	400	400	400	...	8.000
Jember	Garam yodium jenis Halus	3.000	3.000	3.000	...	8.000
Jember	Gula Pasir Dalam Negeri	9.700	9.700	9.700	...	12.000
Jember	Ikan Asin Teri	50.000	50.000	50.000	...	86.667
Jember	Jagung Pipilan Kering	4.000	4.000	4.000	...	8.000
Jember	Kacang Hijau	16.000	16.000	17.000	...	24.000
Jember	Kacang Kedelai Eks Impor	9.700	9.700	9.700	...	9.000
Jember	Kacang Tanah	16.000	16.000	17.000	...	26.000
Jember	Kentang	9.000	9.000	9.000	...	12.000
Jember	Ketela Pohon	1.500	1.500	1.500	...	3.667
Jember	Kol / Kubis	4.000	4.000	3.000	...	11.000
Jember	Mie Instant Indomie Rasa Kari Ayam	1.500	1.500	1.500	...	2.400
Jember	Minyak Curah	11.200	11.200	11.500	...	12.000

Tabel 4.1 (Lanjutan)

DAERAH	KOMODITAS	01/01/2014	01/02/2014	01/03/2014	...	31/12/2020
Jember	Minyak Goreng Bimoli Botol/Kemasan 2 ltr	26.000	26.000	26.000	...	26.000
Jember	Susu Bubuk Merk Bendera (Instant)	27.000	27.000	27.000	...	40.000
Jember	Susu Bubuk Merk Indomilk (Instant)	30.000	30.000	30.000	...	38.000
Jember	Susu Kental Manis Merk Bendera	8.000	8.000	8.000	...	9.850
Jember	Susu Kental Manis Merk Indomilk	8.000	8.000	8.000	...	9.833
Jember	Telur Ayam Kampung	16.500	16.500	16.500	...	37.667
Jember	Telur Ayam Ras/Petelur	15.500	15.500	16.500	...	23.333
Jember	Tepung Terigu Segitiga Biru (Kw Medium)	6.500	6.500	6.500	...	9.000
Jember	Tomat	14.000	14.000	14.000	...	9.333
Jember	Wortel	6.000	6.000	5.000	...	10.333

b. Data Indeks Harga Konsumen (IHK)

Data yang diambil dari BPS merupakan data Indeks Harga Konsumen yang diterbitkan per tahun oleh BPS. Data IHK terbitan BPS disebarikan dalam file beformat PDF dan bisa diunduh secara bebas di situs webnya. Total data yang didapat adalah 960 data, karena data IHK yang diberikan merupakan data bulanan yang berasal dari 8 kabupaten/kota. Data IHK dirangkum dalam Tabel 4.2, Tabel 4.3, Tabel 4.4, Tabel 4.5, Tabel 4.6, Tabel 4.7, Tabel 4.8, dan Tabel 4.9.

Tabel 4. 2 Data IHK Kabupaten Jember

	2014	2015	2016	2017	2018	2019	2020
JANUARI	164,80	177,70	124,94	131,88	133,04	136,18	104,11
FEBRUARI	168,53	122,28	127,45	129,79	133,51	134,77	105,48
MARET	172,92	120,60	127,83	128,24	131,76	133,91	104,50
APRIL	171,31	119,92	126,96	127,56	132,23	135,82	103,43
MEI	167,62	117,81	126,27	128,75	132,74	137,45	102,67
JUNI	168,50	118,63	125,81	128,59	135,68	135,76	103,10

Tabel 4.2 (Lanjutan)

	2014	2015	2016	2017	2018	2019	2020
JULI	177,12	118,67	126,77	128,37	134,10	136,89	102,71
AGUSTUS	179,85	122,26	128,13	127,57	133,15	136,70	101,33
SEPTEMBER	177,82	123,70	126,38	127,46	131,44	134,47	101,07
OKTOBER	175,12	123,67	126,15	125,93	131,59	134,72	101,17
NOVEMBER	173,15	122,85	124,40	127,46	132,81	135,87	102,96
DESEMBER	177,70	123,84	125,48	130,63	135,51	138,26	104,56

Tabel 4.3 Data IHK Kabupaten Banyuwangi

	2014	2015	2016	2017	2018	2019	2020
JANUARI		129,60	131,90	130,51	126,25	123,64	98,92
FEBRUARI		125,88	131,76	128,42	126,61	122,84	98,51
MARET		126,25	130,77	126,27	126,74	123,35	98,79
APRIL		123,94	129,38	125,46	125,46	123,83	98,20
MEI		125,60	129,17	125,53	125,40	124,21	98,88
JUNI		126,42	131,00	124,44	127,09	124,30	99,47
JULI		127,34	131,22	124,76	126,11	124,54	97,66
AGUSTUS		128,36	129,61	123,08	124,78	123,41	95,84
SEPTEMBER		127,52	128,96	122,03	121,23	121,83	95,28
OKTOBER		126,15	129,01	120,11	119,85	121,46	95,90
NOVEMBER		126,43	129,45	121,20	120,40	122,09	97,54
DESEMBER		129,49	130,69	123,83	122,59	122,52	99,62

Tabel 4.4 Data IHK Kabupaten Sumenep

	2014	2015	2016	2017	2018	2019	2020
JANUARI	169,53	120,91	129,94	128,73	133,19	136,60	107,62
FEBRUARI	171,95	119,07	129,03	129,00	132,49	134,43	107,82
MARET	175,72	119,43	126,47	127,12	130,86	133,68	107,02
APRIL	170,81	117,96	125,43	125,93	129,92	135,60	106,27
MEI	167,87	119,42	126,42	128,31	131,16	138,55	105,82
JUNI	170,19	120,40	128,31	128,30	134,02	138,44	105,37
JULI	183,62	123,76	130,07	128,25	133,18	137,85	103,95
AGUSTUS	184,80	123,29	128,09	126,18	131,43	137,46	101,89
SEPTEMBER	175,73	122,25	127,04	125,59	131,29	135,84	101,32
OKTOBER	177,66	122,39	126,54	125,21	131,68	136,98	100,41
NOVEMBER	175,05	123,97	128,31	127,74	132,52	139,23	104,10
DESEMBER	176,27	127,56	131,00	130,48	134,99	140,72	107,01

Tabel 4. 5 Data IHK Kota Kediri

	2014	2015	2016	2017	2018	2019	2020
JANUARI	160,46	119,12	121,73	121,12	121,87	126,89	106,78
FEBRUARI	164,21	116,22	120,17	122,12	122,88	126,35	108,67
MARET	166,24	115,21	120,85	120,23	123,05	127,50	107,61
APRIL	164,90	113,60	120,10	119,99	123,50	129,80	106,51
MEI	161,74	114,38	119,91	121,40	122,42	128,93	105,55
JUNI	162,25	115,11	120,31	121,55	123,52	128,33	106,54
JULI	169,87	116,35	122,65	120,91	123,89	130,19	105,13
AGUSTUS	171,87	116,56	120,25	119,88	122,68	127,98	104,11
SEPTEMBER	168,76	115,70	119,24	118,83	121,62	125,85	104,16
OKTOBER	166,10	114,45	118,29	118,09	121,86	126,66	103,83
NOVEMBER	166,81	114,92	120,70	119,39	123,34	128,74	106,19
DESEMBER	167,63	118,76	121,68	121,72	125,45	130,87	107,90

Tabel 4. 6 Data IHK Kota Malang

	2014	2015	2016	2017	2018	2019	2020
JANUARI	172,30	124,75	134,29	140,61	144,04	144,90	103,35
FEBRUARI	176,85	123,13	135,14	141,70	143,87	142,62	104,43
MARET	181,78	122,22	136,01	139,67	143,54	142,94	103,78
APRIL	179,65	121,33	134,63	138,53	144,08	143,99	103,05
MEI	176,56	122,78	133,98	140,04	141,84	145,76	102,95
JUNI	178,39	124,15	135,55	139,09	141,92	145,05	104,08
JULI	189,76	125,24	138,20	138,87	143,39	147,97	104,00
AGUSTUS	191,26	126,64	136,52	136,54	141,36	148,36	103,73
SEPTEMBER	185,85	125,70	136,17	135,91	139,70	147,24	103,45
OKTOBER	184,54	125,31	134,85	136,01	140,49	145,92	102,90
NOVEMBER	184,71	125,93	136,98	137,09	141,17	146,19	104,37
DESEMBER	187,76	130,85	137,32	140,32	143,27	147,67	105,85

Tabel 4. 7 Data IHK Kota Probolinggo

	2014	2015	2016	2017	2018	2019	2020
JANUARI	156,47	123,23	126,41	125,53	126,44	130,13	104,90
FEBRUARI	159,19	122,53	125,61	124,58	126,85	129,29	105,53
MARET	163,11	120,49	124,53	122,56	125,46	128,21	104,66
APRIL	157,98	118,92	124,80	122,36	125,77	129,59	104,31
MEI	157,37	120,30	124,74	123,46	126,31	131,54	104,18
JUNI	160,21	121,24	125,45	123,62	128,13	132,14	104,97
JULI	168,41	122,64	126,45	123,18	128,54	132,74	104,54

Tabel 4.7 (Lanjutan)

	2014	2015	2016	2017	2018	2019	2020
AGUSTUS	172,11	123,30	125,30	122,33	127,13	133,81	102,67
SEPTEMBER	167,12	122,89	124,16	122,56	125,52	132,40	101,09
OKTOBER	165,52	122,42	122,73	121,61	125,66	132,23	101,62
NOVEMBER	165,64	122,78	123,88	122,70	127,16	133,47	103,80
DESEMBER	168,89	124,25	124,82	125,55	129,46	134,31	106,30

Tabel 4. 8 Data IHK Kota Madiun

	2014	2015	2016	2017	2018	2019	2020
JANUARI	166,39	120,85	124,37	125,09	130,23	131,92	105,45
FEBRUARI	169,70	119,45	122,92	125,33	131,19	131,09	107,30
MARET	175,56	119,13	123,35	133,38	130,39	131,94	107,29
APRIL	171,82	118,01	124,57	122,84	130,49	134,06	105,35
MEI	165,44	118,61	124,03	124,77	130,16	135,67	104,42
JUNI	166,64	119,59	124,56	125,55	132,13	134,61	104,14
JULI	180,17	121,61	126,60	124,75	132,17	135,23	102,90
AGUSTUS	182,34	121,94	122,76	123,57	130,13	134,05	101,82
SEPTEMBER	175,56	119,92	121,59	123,39	127,95	131,80	101,38
OKTOBER	175,58	119,14	120,41	123,04	127,90	131,79	101,69
NOVEMBER	172,98	119,10	122,64	123,64	129,07	132,16	103,79
DESEMBER	174,61	121,39	123,89	126,45	130,34	133,55	105,61

Tabel 4. 9 Data IHK Kota Surabaya

	2014	2015	2016	2017	2018	2019	2020
JANUARI	170,78	126,69	133,24	135,97	138,75	142,92	104,41
FEBRUARI	175,88	125,21	132,80	135,47	139,02	141,64	105,60
MARET	180,64	125,01	133,01	134,15	138,92	141,31	105,68
APRIL	178,30	124,09	133,71	133,24	139,38	143,91	105,18
MEI	176,41	124,57	132,90	134,68	139,45	144,56	105,05
JUNI	176,52	125,84	135,02	135,27	140,43	143,83	105,06
JULI	186,30	126,49	137,04	135,07	141,06	144,60	103,43
AGUSTUS	187,80	128,67	135,89	132,72	139,78	143,38	102,60
SEPTEMBER	184,08	128,37	134,23	131,83	139,07	141,84	102,11
OKTOBER	181,71	126,47	132,20	130,99	138,59	141,15	101,90
NOVEMBER	181,77	126,26	132,82	131,49	138,73	142,45	103,02
DESEMBER	184,65	130,12	133,98	135,44	141,69	145,22	104,61

c. Nilai IHK dengan Rumus Laspeyres

Rumus yang digunakan dalam penghitungan nilai IHK oleh BPS adalah Rumus Laspeyres sebagaimana tertulis pada Persamaan 2.1. Data yang dibutuhkan dalam penghitungan IHK adalah data harga komoditas dan kuantitas barang yang dikonsumsi oleh masyarakat. Data harga komoditas dapat diambil dari Website Siskaperbapo, sedangkan data kuantitas dapat diperoleh di Website BPS Provinsi Jawa Timur.

Data harga konsumen yang digunakan dimulai pada tahun 2014-2020. Tahun 2014 data harga konsumen didasarkan pada Survey Biaya Hidup (SBH) Tahun 2012. Januari 2014 merupakan tahun pertama diberlakukan indeks 2012=100 untuk menggantikan indeks sebelumnya yaitu 2007=100. Hal ini dikarenakan SBH terakhir sebelum tahun 2014 dilakukan pada tahun 2012.

Pada penelitian ini data IHK yang digunakan berasal dari data IHK yang sudah diolah berdasarkan publikasi yang diunduh dari situs jatim.bps.go.id. Contoh penghitungan IHK pada Tabel 4.10.

Tabel 4. 10 Tabel Variabel Penentu Nilai IHK

Nama Barang	Harga (P)		Kuantitas (Q)	
	2017	2019	2017	2019
Padi-padian	55229	59459	111,58	103,62
Daging	22463	20866	17,77	19,8
Telur dan Susu	26981	28535	18,44	20,94
Kacang-kacangan	15328	15433	20,84	48,27

Variabel yang digunakan untuk menentukan nilai IHK menggunakan rumus Laspeyres adalah harga komoditas dan kuantitas konsumsi yang digunakan pada

suatu periode. Variabel harga dilambangkan dengan P dan kuantitas dilambangkan dengan Q . Variabel P yang digunakan terbagi menjadi P_n (harga tahun berjalan) dan P_0 (harga tahun dasar, biasanya tahun paling awal yang digunakan sebagai acuan).

Jadi misalnya kita akan mencari nilai IHK untuk beberapa komoditas makanan di tahun 2019 dengan tahun dasar 2017. Maka perhitungannya akan menjadi seperti berikut:

$$\begin{aligned}
 I_n &= \frac{\sum(P_n \times Q_0)}{\sum(P_0 \times Q_0)} \times 100 \\
 &= \frac{(59459 \times 111,58) + (20866 \times 17,77) + (28535 \times 18,44) + (15433 \times 20,84)}{(55229 \times 111,58) + (22463 \times 17,77) + (26981 \times 18,44) + (15328 \times 20,84)} \times 100 \\
 &= \frac{7853033,16}{7378584,49} \times 100 \\
 &= 106,43
 \end{aligned}$$

Berdasarkan hasil contoh perhitungan nilai IHK menggunakan rumus Laspeyres ditemukan hasil IHK sebesar 106,43. Ketika nilai IHK < 100 , maka tingkat harga konsumen/harga ecer lebih rendah daripada tahun dasar. Sedangkan jika nilai IHK > 100 , maka tingkat harga konsumen lebih tinggi daripada tahun dasar (Statistik, 2023). Itu artinya, pada waktu tersebut terdapat kenaikan rata-rata sebesar 6,43 kali lipat dibandingkan waktu sebelumnya.

d. Impor Data dalam Python

Data yang sudah dikumpulkan selanjutnya mulai diolah ke dalam Bahasa Pemrograman Python menggunakan Jupyter Notebook. Sebelum mulai melakukan proses impor data, terlebih dahulu dilakukan impor *packages* yang digunakan

dalam proses implementasi model. *Packages* yang digunakan antara lain: numpy, pandas, sklearn.svm untuk impor SVR, sklearn.preprocessing untuk impor PR, model_selection untuk GridSearchCV, metrics, matplotlib, dan seaborn (Gambar 4. 1). Setelah prose impor *packages* selesai dilanjutkan dengan impor *dataset* berformat Microsoft Office Excel dengan nama *file* Dataset.xlsx (Gambar 4. 2). Hasil impor *dataset* ada pada Gambar 4. 3 dan Gambar 4. 4.

```
In [41]: import numpy as np
import pandas as pd
from sklearn.metrics import r2_score
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt
import seaborn as sns
```

Gambar 4. 1 Impor Packages Python

```
In [42]: df = pd.read_excel('Dataset.xlsx', sheet_name='Data RealEstate')
df.head()
Out[42]:
```

Gambar 4. 2 Impor Dataset

```
In [43]: df.head(5)
```

Index	Tanggal	Grupp Memb	Bangsa Pilih	Dinas Sampunan	Dinas Pilih	Dinas Memb	Bangsa	Catun Memb	Catun Pilih	Total Memb Memb	Total Memb Pilih	Total Memb Memb	Total Memb Pilih	Total Memb Memb	Total Memb Pilih
0	2014-01-01	20000	11000	8000	7000	9000	7000	8000	20000	10000	8000	9000	10000	10000	10000
1	2014-01-02	20000	11000	8000	7000	9000	7000	8000	20000	10000	8000	9000	10000	10000	10000
2	2014-01-03	20000	11000	8000	7000	9000	7000	8000	20000	10000	8000	9000	10000	10000	10000
3	2014-01-04	20000	11000	8000	7000	9000	7000	8000	20000	10000	8000	9000	10000	10000	10000
4	2014-01-05	20000	11000	8000	7000	9000	7000	8000	20000	10000	8000	9000	10000	10000	10000

3 rows × 15 columns

Gambar 4. 3 Hasil Data yang Sudah Diimpor

```
In [47]: df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 20455 entries, 0 to 20452
Data columns (total 37 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Daerah                20455 non-null  object
 1   Tanggal               20455 non-null  datetime64[ns]
 2   Ruang Pasar           20450 non-null  float64
 3   Ruang Pemas          20450 non-null  float64
 4   Beras Bergenan       20450 non-null  float64
 5   Beras IR34           20450 non-null  float64
 6   Beras Hestik         20450 non-null  float64
 7   Buncis                20450 non-null  float64
 8   Cabe Besar Sissa     20450 non-null  float64
 9   Cabe Runcit          20450 non-null  float64
10   Daging Ayam Broiler  20450 non-null  float64
11   Daging Ayam Kampung 20450 non-null  float64
12   Daging Sapi Puncak   20450 non-null  float64
13   Garam yodium jenis Bata 20450 non-null  float64
14   Garam yodium jenis Halus 20450 non-null  float64
15   Dulu Pasir Dalam Negeri 20450 non-null  float64
16   Ikan Asin Teri       20450 non-null  float64
17   Jagung Pislak Haring 20450 non-null  float64
18   Kacang Hijau         20450 non-null  float64
19   Kacang Kacil Eka Iner 20450 non-null  float64
20   Kacang Tanah         20450 non-null  float64
21   Jering               20450 non-null  float64
22   Ketela Pohon         20450 non-null  float64
23   Kac 7 Kubis         20450 non-null  float64
24   Nila Tawar Indoneia Rasi Kori byan 20450 non-null  float64
25   Niyah Gurah         20450 non-null  float64
26   Niyah Goreng Biscit Sotol/Kewasan 2 Itr 20455 non-null  float64
27   Susu Bubuk Neri Bendera (Instan) 20450 non-null  float64
28   Susu Bubuk Neri Indoniah (Instan) 20450 non-null  float64
29   Susu Kental Pania Neri Bendera 20450 non-null  float64
30   Susu Kental Pania Neri Indoneia 20450 non-null  float64
31   Telur Ayam Kampung 20450 non-null  float64
32   Telur Ayam Ras/Petalar 20450 non-null  float64
33   Tepung Terigu Segitiga Biru (No Reduse) 20450 non-null  float64
34   Tomat                20450 non-null  float64
35   Urtak                20450 non-null  float64
36   UHC                  20455 non-null  float64

dtypes: datetime64[ns](1), float64(36), object(1)
memory usage: 3.0+ MB
```

Gambar 4. 4 Hasil Info Data yang Sudah Diimpor

4.1.2. Praproses Data

Agar data dapat diproses dengan baik menggunakan model SVR dan PR, maka data harus melalui praproses terlebih dahulu. Praproses data terdiri dari penghapusan *missing value* dan integrasi data.

a. Penghapusan *Missing Value*

Penghapusan *missing value* dilakukan pada data yang memiliki nilai 0. Kadang kala terdapat harga komoditas makanan maupun IHK yang memiliki nilai

0. Maka untuk menghilangkan *missing value* tersebut, dilakukan penghapusan pada data. Sehingga dari total 20.453 yang dikumpulkan, tersisa sebanyak 19.723 data (Gambar 4. 5).

```
In [40]: # drop missing values
df2 = df.dropna()

In [41]: df2.shape
Out[41]: (19723, 8)
```

Gambar 4. 5 Proses Penghapusan Missing Value

b. Integrasi Data

Selain melalui proses penghapusan *missing value*, data juga melalui proses integrasi data. Proses ini dilakukan untuk menggabungkan data harga komoditas makanan dengan data IHK-nya. Serta menggabungkan data dari ke 8 kabupaten/kota ke dalam satu dokumen data. Contoh hasil integrasi data dapat dilihat pada Tabel 4.11.

Tabel 4. 11 Contoh Data Hasil Integrasi Harga Komoditas Makanan dan IHK

Daerah	Tanggal	Bawang Merah	Bawang Putih	Beras Bengawan	Beras IR64	...	IHK
Jember	01/01/2014	33.000	11.000	8.500	7.500	...	164,80
Jember	02/01/2014	33.000	11.000	8.500	7.500	...	164,80
Jember	03/01/2014	33.000	10.000	8.500	7.500	...	164,80
Jember	04/01/2014	33.000	10.000	8.500	7.500	...	164,80
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Surabaya	12/31/2020	27.600	23.400	12.620	9.450	...	104,61

4.1.3. Pemisahan Data menjadi Data Latih dan Data Uji

Untuk mendapatkan hasil model yang baik, perlu dilakukan proses pembagian data menjadi data latih dan data uji. Data latih digunakan untuk melatih

data, sedangkan data uji digunakan untuk mengevaluasi model. Pada penelitian ini, implementasi model untuk SVR akan dilakukan dengan beberapa eksperimen melalui pembagian data latih sebesar 70%, 80%, dan 90%. Diikuti dengan pembagian data uji untuk sisa datanya sebesar 30%, 20%, dan 10%. Eksperimen dilakukan beberapa kali untuk mengetahui perbedaan hasil model demi mencapai hasil model yang paling baik.

Beberapa eksperimen terdahulu dilakukan dengan memberikan perbandingan yang besar pada data latih daripada data uji. Hal ini dikarenakan data latih yang tinggi dipercaya dapat memberikan varians yang rendah sehingga menghindari adanya *overfitting data*. Jika *overfitting* dapat dihindari, maka model yang dihasilkan akan menjadi baik (Afifah, 2022). Pada penelitian yang dilakukan oleh Dewi, dkk. yang melakukan eksperimen dengan perbandingan data latih 40%, 50%, dan 60%, dan menghasilkan model dengan akurasi tinggi untuk data latih yang lebih besar (Dewi, Kartikasari and Mursityo, 2014). Proporsi perbandingan data yang digunakan dapat dilihat pada Tabel 4. 12.

Tabel 4. 12 Proporsi Perbandingan Data Latih dan Data Uji

	90%:10%	80%:20%	70%:30%
Data Latih	17750	15778	13806
Data Uji	1973	3945	5917

4.2. Pelaksanaan Aksi

Sub bab Pelaksanaan Aksi berisi proses implementasi model untuk menganalisa korelasi dan prediksi nilai IHK. Analisa korelasi dilakukan dengan Pearson Product Moments. Sementara prediksi nilai IHK dilakukan dengan metode Support Vector Regression dan Polynomial Regression.

4.2.1. Analisa Korelasi menggunakan Pearson Product Moments

Untuk mencari keterkaitan antara variabel bebas yaitu harga komoditas makanan dengan variabel terikatnya (IHK), dilakukan analisa korelasi. Metode yang digunakan adalah Pearson Product Moments. Langkah analisa korelasi menggunakan Pearson Product Moments dijelaskan sebagai berikut:

1. Import Packages

Langkah pertama adalah *import packages Python* yang dibutuhkan. Proses *import packages* dapat dilihat pada Gambar 4. 6.

```
In [1]: import numpy as np
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics.pairwise import euclidean_distances
from sklearn.metrics.pairwise import euclidean_distances
import matplotlib.pyplot as plt
import seaborn as sns
```

Gambar 4. 6 Import Packages

2. Import Dataset

Langkah selanjutnya adalah *import dataset* yang berisi data harga komoditas dan IHK. Proses *input dataset* dapat dilihat pada Gambar 4. 7

```
In [2]: df = pd.read_excel('data/ihk', sheet_name='Data Analisis')
# # Create feature matrix
X = pd.DataFrame(df[['komoditas', 'harga', 'ihk']])
```

Gambar 4. 7 Import Dataset

3. Penentuan Variabel X dan Variabel Y

Setelah *dataset* dimasukkan, selanjutnya ditentukan atribut yang termasuk variabel *x* dan variabel *y*. Variabel *x* merupakan variabel penentu, sedangkan variabel *y* merupakan variabel target (Gambar 4. 8).

Variabel *x* terdiri dari 34 komoditas makanan, diantaranya: Bawang Merah, Bawang Putih, Beras Bengawan, Beras IR64, Beras Mentik, Buncis, Cabe Besar Biasa, Cabe Rawit, Daging Ayam Broiler, Daging Ayam Kampung, Daging Sapi Murni, Garam yodium jenis Bata, Garam yodium jenis Halus, Gula Pasir Dalam Negeri, Ikan Asin Teri, Jagung Pipilan Kering, Kacang Hijau, Kacang Kedelai Eks Impor, Kacang Tanah, Kentang, Ketela Pohon, Kol / Kubis, Mie Instant Indomie Rasa Kari Ayam, Minyak Curah, Minyak Goreng Bimoli Botol/Kemasan 2 ltr, Susu Bubuk Merk Bendera (Instant), Susu Bubuk Merk Indomilk (Instant), Susu Kental Manis Merk Bendera, Susu Kental Manis Merk Indomilk, Telur Ayam Kampung, Telur, Ayam Ras/Petelur, Tepung Terigu Segitiga Biru (Kw Medium), Tomat, dan Wortel. Sedangkan variabel *y* adalah IHK.



Gambar 4. 8 Proses Penentuan Variabel X dan Variabel Y

4. Plot Diagram Pencar untuk Visualisasi Korelasi menggunakan Pearson Product Moment

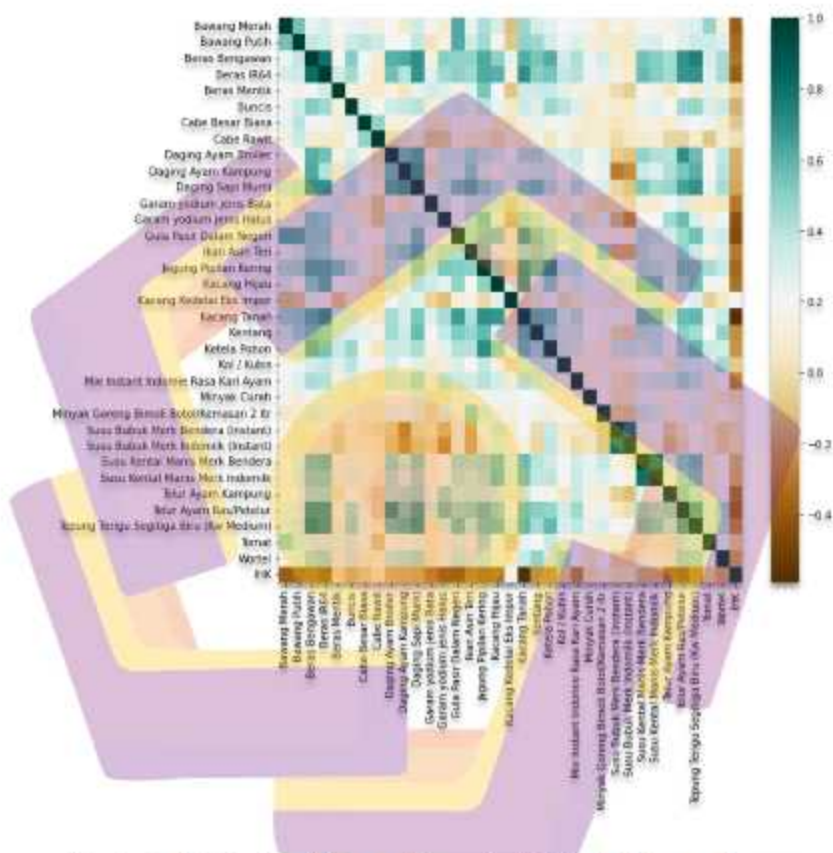
Hal ini dilakukan untuk mengetahui hubungan yang terbentuk antara variabel melalui diagram agar dapat dilihat secara visual (sGambar 4. 9 dan Gambar 4. 10).

```

R> plot(FigureCaption(4),10)
G = as.comparison(pearson)
M = vis.highlight(1000,0.01)

```

Gambar 4. 9 Plot Diagram Pencar



Gambar 4. 10 Visualisasi Diagram Pencar Analisa Korelasi dengan Pearson Product Moment

5. Tampilkan Data Hasil Korelasi

Selain ditampilkan dalam format gambar, dicari juga data hasil korelasi melalui data tabel (Gambar 4. 11 dan Tabel 4. 13).

Gambar 4. 11 Proses Penampilan Data Hasil Korelasi

Tabel 4. 13 Data Hasil Korelasi

	IHK		IHK		IHK
Bawang Merah	-0.385511	Garam yodium jenis Halus	-0.422824	Minyak Goreng Bimoli Botol/Kemasan 2 ltr	0.078829
Bawang Putih	-0.338958	Gula Pasir Dalam Negeri	-0.452802	Susu Bubuk Merk Bendera (Instant)	-0.000103
Beras Bengawan	-0.315241	Ikan Asin Teri	-0.300677	Susu Bubuk Merk Indomilk (Instant)	0.021608
Beras IR64	-0.377965	Jagung Pipilan Kering	-0.352298	Susu Kental Manis Merk Bendera	-0.150037
Beras Mentik	-0.033197	Kacang Hijau	-0.323250	Susu Kental Manis Merk Indomilk	-0.155314
Buncis	-0.152444	Kacang Kedelai Eks Impor	0.224138	Telur Ayam Kampung	-0.326240
Cabe Besar Biasa	-0.077660	Kacang Tanah	-0.593384	Telur Ayam Ras/Petelur	-0.386049
Cabe Rawit	0.022240	Kentang	-0.241689	Tepung Terigu Segitiga Biru (Kw Medium)	-0.263756
Daging Ayam Broiler	-0.155022	Ketela Pohon	-0.231153	Tomat	-0.074656
Daging Ayam Kampung	-0.168504	Kol / Kubis	-0.078749	Wortel	-0.189871
Daging Sapi Murni	-0.178500	Mie Instant Indomie Rasa Kari Ayam	-0.265385		
Garam yodium jenis Bata	-0.305084	Minyak Curah	0.006631		

Berdasarkan hasil visualisasi diagram pada Gambar 4. 10, dapat disimpulkan bahwa hubungan yang terjadi adalah korelasi negatif yang sempurna antara variabel x dan y . Artinya semakin tinggi atau naik nilai x , maka semakin turun nilai y . Begitu pula sebaliknya. Hal ini menunjukkan bahwa semakin tinggi harga komoditas makanan, maka hal tersebut merupakan indikasi adanya inflasi, sehingga dapat menurunkan harga beli masyarakat yang diikuti dengan menurunnya nilai IHK.

Jika dilihat secara keseluruhan mayoritas komoditas memiliki hubungan yang negatif dengan nilai IHK. Namun tidak seluruh komoditas memiliki hubungan negatif. Terdapat juga beberapa komoditas yang memiliki hubungan positif dengan IHK, misalnya: Cabe Rawit, Kacang Kedelai Eks Impor, Minyak Curah, Minyak Goreng Bimoli Botol/Kemasan 2 ltr, dan Susu Bubuk Merk Indomilk (Instant). Hal ini menunjukkan adanya hubungan yang positif, sehingga semakin naik nilai variabel x , maka semakin naik pula variabel y -nya. Itu berarti naik atau turun harga komoditas tersebut, tidak memengaruhi daya beli masyarakat terhadap komoditasnya.

Data harga komoditas dengan nilai IHK memiliki korelasi berdasarkan analisa korelasi menggunakan Pearson Product Moment, namun korelasi yang terbentuk terbilang rendah. Korelasi yang terjadi antar variabel berkisar pada rentang 0.000103 hingga 0.593384. Rentang ini berada pada tingkat korelasi sangat, rendah, rendah, dan sedang. Hasil tingkatan korelasi secara lengkap disajikan dalam Tabel 4. 14.

Tabel 4. 14 Tingkat Korelasi Variabel X dan Variabel Y

Komoditas	IHK	Tingkat Korelasi
Bawang Merah	-0.385511	Korelasi Rendah
Bawang Putih	-0.338958	Korelasi Rendah
Beras Bengawan	-0.315241	Korelasi Rendah
Beras IR64	-0.377965	Korelasi Rendah
Beras Mentik	-0.033197	Korelasi Sangat Rendah
Buncis	-0.152444	Korelasi Sangat Rendah
Cabe Besar Biasa	-0.077660	Korelasi Sangat Rendah
Cabe Rawit	0.022240	Korelasi Sangat Rendah
Daging Ayam Broiler	-0.155022	Korelasi Sangat Rendah
Daging Ayam Kampung	-0.168504	Korelasi Sangat Rendah
Daging Sapi Murni	-0.178500	Korelasi Sangat Rendah
Garam yodium jenis Bata	-0.305084	Korelasi Rendah
Garam yodium jenis Halus	-0.422824	Korelasi Sedang
Gula Pasir Dalam Negeri	-0.452802	Korelasi Sedang
Ikan Asin Teri	-0.300677	Korelasi Rendah
Jagung Pipilan Kering	-0.352298	Korelasi Rendah
Kacang Hijau	-0.323250	Korelasi Rendah
Kacang Kedelai Eks Impor	0.224138	Korelasi Rendah
Kacang Tanah	-0.593384	Korelasi Sedang
Kentang	-0.241689	Korelasi Rendah
Ketela Pohon	-0.231153	Korelasi Rendah
Kol / Kubis	-0.078749	Korelasi Sangat Rendah
Mie Instant Indomie Rasa Kari Ayam	-0.265385	Korelasi Rendah
Minyak Curah	0.006631	Korelasi Sangat Rendah
Minyak Goreng Bimoli Botol/Kemasan 2 ltr	0.078829	Korelasi Sangat Rendah
Susu Bubuk Merk Bendera (Instant)	-0.000103	Korelasi Sangat Rendah
Susu Bubuk Merk Indomilk (Instant)	0.021608	Korelasi Sangat Rendah
Susu Kental Manis Merk Bendera	-0.150037	Korelasi Sangat Rendah
Susu Kental Manis Merk Indomilk	-0.155314	Korelasi Sangat Rendah
Telur Ayam Kampung	-0.326240	Korelasi Rendah
Telur Ayam Ras/Petelur	-0.386049	Korelasi Rendah
Tepung Terigu Segitiga Biru (Kw Medium)	-0.263756	Korelasi Rendah
Tomat	-0.074656	Korelasi Sangat Rendah
Wortel	-0.189871	Korelasi Sangat Rendah

4.2.2. Prediksi Nilai IHK menggunakan Support Vector Regression

Proses prediksi menggunakan SVR akan dilakukan sebanyak tiga kali dengan perbandingan data latih dan data uji yang berbeda. Perbandingan yang digunakan antara data latih dan data ujinya adalah 70%:30%, 80%:20%, dan 90%:10%. Hal ini dilakukan untuk mengetahui perbedaan hasil model demi memperkaya hasil eksperimennya.

Kernel yang digunakan adalah Radial Basis Function (RBF). Prediksi menggunakan kernel RBF membutuhkan tiga *hyperparameter*, diantaranya: Cost (C), Gamma, dan Epsilon. Pengaturan nilai Cost diatur pada rentang nilai $C \in \{0.1, 0.5, 1.0, 1.5, 2.0\}$. Sedangkan Gamma diatur pada nilai $\text{Gamma} \in \left\{ \text{scale} = \frac{1}{p\sigma^2}, \text{auto} = \frac{1}{p} \right\}$. Dimana p merujuk pada jumlah variabel, sedangkan σ^2 adalah ragam variabel independen yang digunakan. Proses pencarian *hyperparameter* tersebut dilakukan menggunakan metode Grid Search Optimization dengan K-Fold Cross Validation. Nilai k yang digunakan adalah $k=5$. Pencarian *hyperparameter* juga dilakukan sebanyak tiga kali sesuai dengan perbandingan data latih dan data uji yang digunakan.

Langkah prediksi nilai IHK menggunakan SVR dijelaskan sebagai berikut:

1. Langkah pertama yang dilakukan adalah mencari *hyperparameter* terbaik menggunakan Grid Search Optimization. Proses pencarian *hyperparameter tuning* pada SVR dengan Kernel RBF dilakukan beberapa kali dengan percobaan beberapa nilai C . Hasil *hyperparameter* terbaik ditunjukkan pada nilai $C = 2$, Gamma = 'scale', dan Epsilon = 1.0 (Gambar 4. 12). Hasil

pencarian *hyperparameter tuning* untuk masing-masing perbandingan data ditunjukkan pada Gambar 4. 13, Gambar 4. 14, Gambar 4. 15, dan Tabel 4. 15;

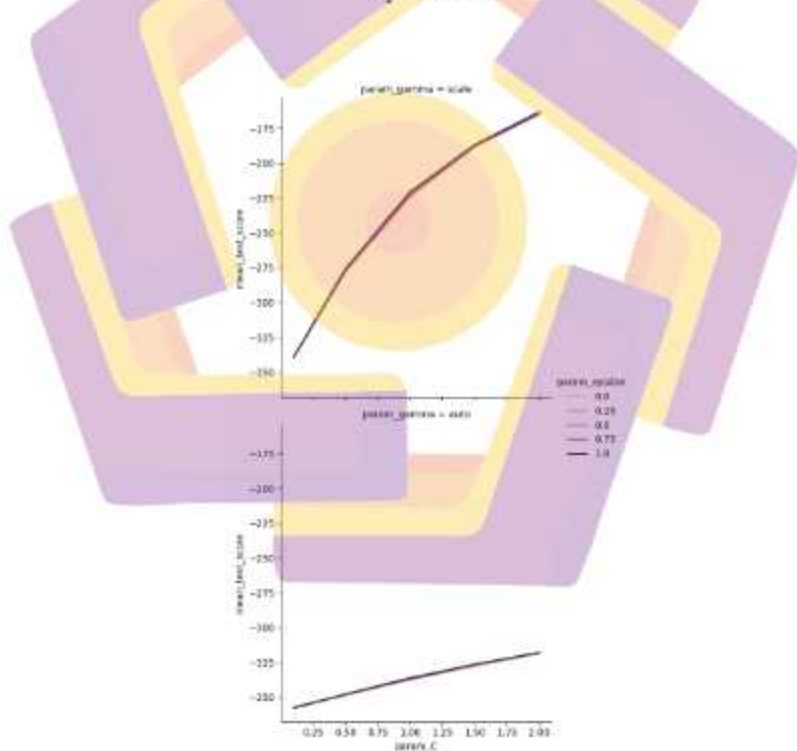
```

In [10]: grid_search = GridSearchCV(
    svm, param_grid, scoring='neg_log_loss',
    cv=5, n_jobs=-1, verbose=1)
grid_search.fit(X_train, y_train)

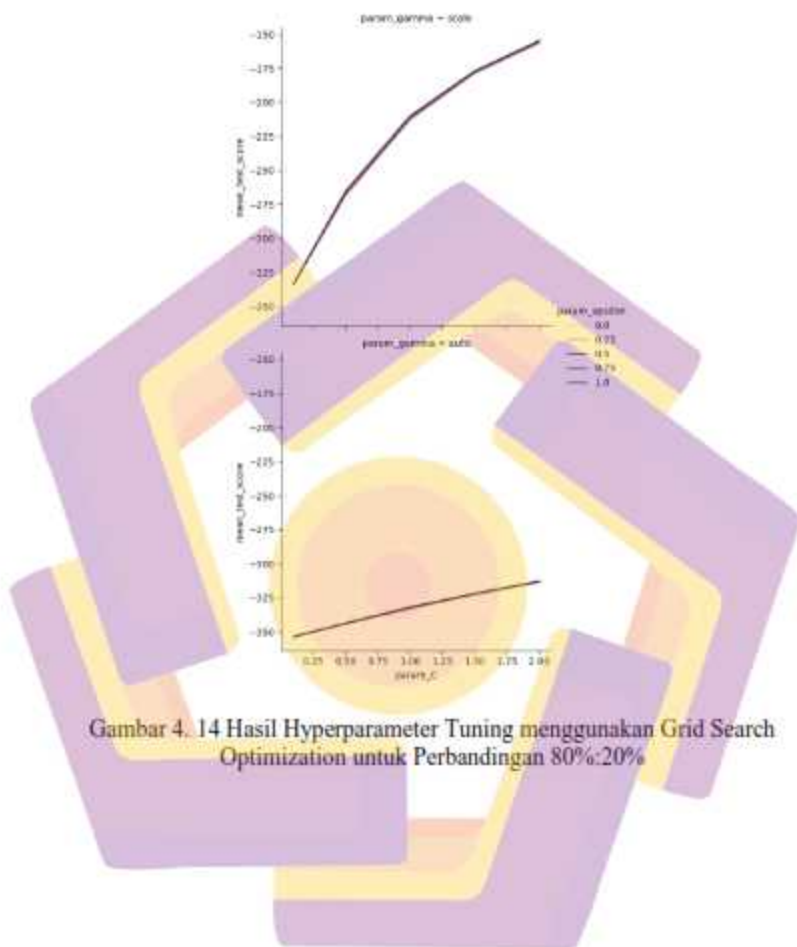
Out[10]: GridSearchCV(cv=5, estimator=SVC(), n_jobs=-1,
    param_grid={'C': [0.1, 0.5, 1.0, 1.5, 2.0],
    'gamma': [0.05, 0.1, 0.2, 0.3, 0.4]},
    scoring='neg_log_loss', verbose=1)
Out[11]: GridSearchCV(cv=5, estimator=SVC(), n_jobs=-1,
    param_grid={'C': [0.1, 0.5, 1.0, 1.5, 2.0],
    'gamma': [0.05, 0.1, 0.2, 0.3, 0.4]},
    scoring='neg_log_loss', verbose=1)

```

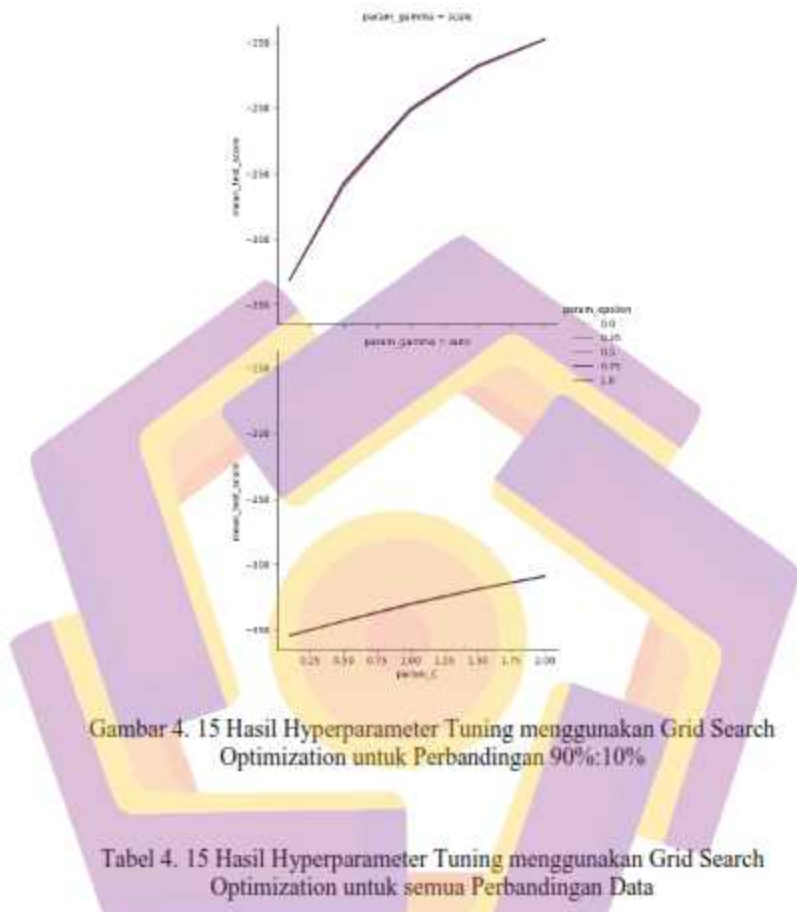
Gambar 4. 12 Proses Pencarian Hyperparameter menggunakan Grid Search Optimization



Gambar 4. 13 Hasil Hyperparameter Tuning menggunakan Grid Search Optimization untuk Perbandingan 70%:30%



Gambar 4. 14 Hasil Hyperparameter Tuning menggunakan Grid Search Optimization untuk Perbandingan 80%:20%



Hyperparameter	70%:30%	80%:20%	90%:10%
C	2.0	2.0	2.0
epsilon	1.0	1.0	1.0
gamma	'scale'	'scale'	'scale'
kernel	'rbf'	'rbf'	'rbf'


```

In [20]: # Metrics of Training Set
print('Metrics of Training Set:')
print('Train set MSE: %s' % str(metrics.mean_squared_error(y_train, y_hat0)) )
print('Train set RMSE: %s' % str(metrics.mean_squared_error(y_train, y_hat0)**0.5) )
print('Train set MAE: %s' % str(metrics.mean_absolute_percentage_error(y_train, y_hat0)) )
print('Training Duration: %s' % str(t1-t0) + " Seconds")

Metrics of Training Set:
Train set MSE: 0.000264320272
Train set RMSE: 0.016258672184
Train set MAE: 0.00018260021898
Training Duration: 0.0118460811 Seconds

```

Gambar 4. 19 Implementasi Model SVR berdasarkan Data Latih dengan perbandingan 80%:20%

```

In [20]: # Metrics of Training Set
print('Metrics of Training Set:')
X = X_train[0:160]
y_train = y_train[0:160]
X_test = X_train[160:]
y_test = y_train[160:]
print('Train set MSE: %s' % str(metrics.mean_squared_error(y_train, y_hat0)) )
print('Train set RMSE: %s' % str(metrics.mean_squared_error(y_train, y_hat0)**0.5) )
print('Train set MAE: %s' % str(metrics.mean_absolute_percentage_error(y_train, y_hat0)) )
print('Training Duration: %s' % str(t1-t0) + " Seconds")

Metrics of Training Set:
Train set MSE: 0.000264320272
Train set RMSE: 0.016258672184
Train set MAE: 0.00018260021898
Training Duration: 0.0118460811 Seconds

```

Gambar 4. 20 Implementasi Model SVR berdasarkan Data Uji dengan perbandingan 80%:20%

5. Percobaan dilakukan kembali dengan perbandingan data latih dan data uji 70%:30% (Gambar 4. 21 dan Gambar 4. 22).

```

In [21]: # Metrics of Training Set
print('Metrics of Training Set:')
X = X_train[0:112]
y_train = y_train[0:112]
X_test = X_train[112:]
y_test = y_train[112:]
print('Train set MSE: %s' % str(metrics.mean_squared_error(y_train, y_hat0)) )
print('Train set RMSE: %s' % str(metrics.mean_squared_error(y_train, y_hat0)**0.5) )
print('Train set MAE: %s' % str(metrics.mean_absolute_percentage_error(y_train, y_hat0)) )
print('Training Duration: %s' % str(t1-t0) + " Seconds")

Metrics of Training Set:
Train set MSE: 0.000264320272
Train set RMSE: 0.016258672184
Train set MAE: 0.00018260021898
Training Duration: 0.0118460811 Seconds

```

Gambar 4. 21 Implementasi Model SVR berdasarkan Data Latih dengan perbandingan 70%:30%

```

In [21]: # Metrics of Training Set
print('Metrics of Training Set:')
print('Train set MSE: %s' % str(metrics.mean_squared_error(y_train, y_hat0)) )
print('Train set RMSE: %s' % str(metrics.mean_squared_error(y_train, y_hat0)**0.5) )
print('Train set MAE: %s' % str(metrics.mean_absolute_percentage_error(y_train, y_hat0)) )
print('Training Duration: %s' % str(t1-t0) + " Seconds")

Metrics of Training Set:
Train set MSE: 0.000264320272
Train set RMSE: 0.016258672184
Train set MAE: 0.00018260021898
Training Duration: 0.0118460811 Seconds

```

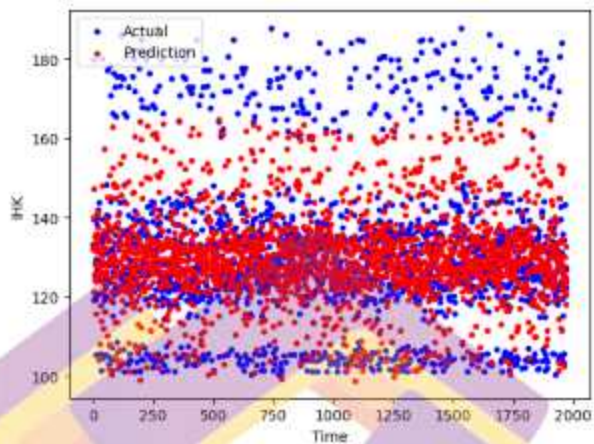
Gambar 4. 22 Implementasi Model SVR berdasarkan Data Uji dengan perbandingan 70%:30%

6. Setelah implementasi model dan evaluasi selesai dilakukan, hasil evaluasi ditampilkan dalam Tabel 4.16. Hasil evaluasi dilakukan menggunakan metode MSE, RMSE, dan MAPE.

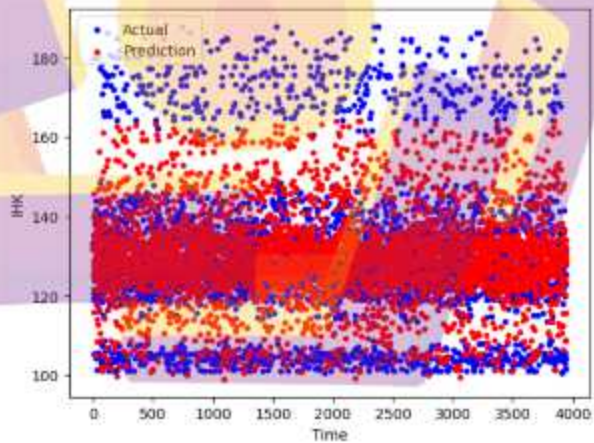
Tabel 4. 16 Hasil Evaluasi SVR dengan Perbandingan Data yang Berbeda

Kriteria	SVR (90%:10%)		SVR (80%:20%)		SVR (70%:30%)	
	Train	Test	Train	Test	Train	Test
MSE	137.025	139.775	141.589	146.540	147.679	148.002
RMSE	11.705	11.822	11.899	12.105	12.152	12.165
MAPE	0.059	0.060	0.060	0.061	0.061	0.061
Compile Time	740.403	5.229	582.651	4.657	443.349	9.167

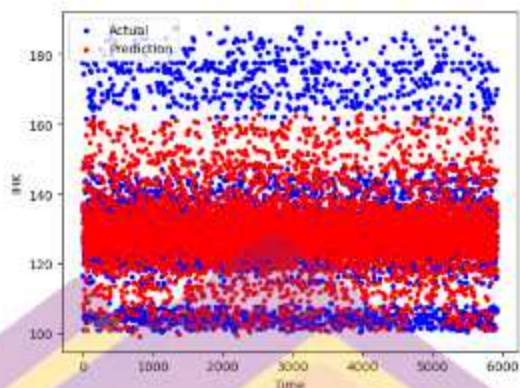
7. Selanjutnya dilakukan plot data untuk menunjukkan perbandingan antara data sebenarnya dan data hasil prediksi. Plot titik berwarna biru melambangkan data sebenarnya, sedangkan titik berwarna merah melambangkan data hasil prediksi. Plot perbandingan data sebenarnya dan hasil prediksi dapat dilihat pada Gambar 4. 23, Gambar 4. 24, dan Gambar 4. 25.



Gambar 4. 23 Plot Data Sebenarnya dan Data Hasil Prediksi menggunakan SVR untuk Perbandingan Data Latih dan Data Uji 90%:10%



Gambar 4. 24 Plot Data Sebenarnya dan Data Hasil Prediksi menggunakan SVR untuk Perbandingan Data Latih dan Data Uji 80%:20%



Gambar 4. 25 Plot Data Sebenarnya dan Data Hasil Prediksi menggunakan SVR untuk Perbandingan Data Latih dan Data Uji 70%:30%

8. Hasil prediksi nilai IHK menggunakan model SVR kemudian dibandingkan dengan data sebenarnya pada publikasi nilai IHK oleh BPS untuk mengetahui tingkat akurasi model. Perbandingan data sebenarnya dengan data hasil prediksi dapat dilihat pada Tabel 4. 17.

Tabel 4. 17 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan SVR

Tempat	Waktu	Data Sebenarnya	Data Hasil Prediksi		
			90%:10%	80%:20%	70%:30%
Jember	Januari 2014	164,8	159,0448	157,5187	156,5632
Jember	Februari 2014	168,53	160,9732	154,3498	158,6616
Jember	Maret 2014	172,92	157,3313	155,8586	155,7209
Jember	April 2014	171,31	154,6495	153,2031	152,9014
Jember	Mei 2014	167,62	153,1426	151,6543	150,7115
Jember	Juni 2014	168,5	150,015	148,5763	147,5757
Jember	Juli 2014	177,12	146,0624	144,3325	143,1529
⋮	⋮	⋮	⋮	⋮	⋮
Surabaya	Desember 2020	104,61	126,8231	122,6174	122,9898

Berdasarkan hasil evaluasi yang ditunjukkan pada Tabel 4. 16, semakin kecil perbandingan data latih yang digunakan, semakin buruk performa modelnya. Nilai MSE yang terbentuk cenderung tinggi, hal ini disesuaikan juga dengan skala data yang digunakan. Semakin banyak data yang digunakan, semakin besar pula nilai MSE yang terbentuk. Nilai MSE yang semakin tinggi menunjukkan nilai kesalahan yang semakin besar. Pada percobaan menggunakan data latih sebesar 90%, nilai MSE yang terbentuk adalah 137,025. Pada percobaan dengan data latih sebesar 80% dan 70%, berturut-turut nilai MSE yang terbentuk adalah 141,589 dan 147,679.

Evaluasi selanjutnya dilakukan menggunakan RMSE. Sama halnya dengan hasil evaluasi menggunakan MSE, hasil yang ditunjukkan mengindikasikan performa yang semakin buruk ketika data latih yang digunakan semakin berkurang atau sedikit. Pada percobaan menggunakan data latih 90%, 80%, dan 70% nilai RMSE yang terbentuk adalah 11,705, 11,899, dan 12,152. Nilai MAPE yang terbentuk pun masih terbilang sama, untuk percobaan data latih menempati nilai 0,059, 0,060, dan 0,061. Berdasarkan hasil evaluasi dengan ketiga metode yang digunakan dapat disimpulkan bahwa ketika data latih yang digunakan semakin banyak, maka performa model semakin baik. Begitu pula sebaliknya.

Dalam segi kecepatan waktu pemrosesan data, SVR memiliki waktu tunggu yang relatif lama. Namun semakin besar persentase data latih yang digunakan, semakin lama proses tunggunya. Proses implementasi model untuk data latih 90% membutuhkan waktu 740,403 sekon, sedangkan sedangkan untuk data uji 5,229 sekon. Sementara untuk percobaan data latih dengan perbandingan 80% dan 70%

membutuhkan waktu 582,651 dan 443,349 sekon. Hal ini berarti semakin kecil persentase data latih yang digunakan maka waktu prosesnya semakin cepat.

4.2.3. Prediksi Nilai IHK menggunakan Polynomial Regression

Implementasi model PR dilakukan dengan menggunakan 3 derajat atau orde yang berbeda, yaitu Orde 1, Orde 2, dan Orde 3. Pelatihan data dilakukan terhadap masing-masing orde yang telah ditentukan. Setelah mendapatkan hasil pelatihan data, kemudian dilakukan evaluasi untuk menentukan model mana yang terbaik menggunakan data uji.

a. Implementasi PR menggunakan Orde 1

Proses implementasi regresi polinomial yang pertama menggunakan derajat 1 atau orde 1 (Gambar 4. 26). Implementasi juga bersamaan dengan proses evaluasi menggunakan MSE, RMSE, dan MAPE (Gambar 4. 27 dan Gambar 4. 28). Setelah itu dilakukan proses visualisasi perbandingan data sebenarnya dan data hasil prediksi yang dapat dilihat pada Tabel 4. 18 dan Gambar 4. 29.

```
In [11]: # Fit the Polynomial regression to the dataset
fit = LinearRegression()
fit.fit(X_train, y_train)
```

Gambar 4. 26 Implementasi Model PR Orde 1

```
In [10]: # Metrics of Training Set
print("Metrics of Training Set")
y_train_min = fit.predict(X_train)
print("Train set MSE: " + str(metrics.mean_squared_error(y_train, y_train_min)))
print("Train set RMSE: " + str(np.sqrt(metrics.mean_squared_error(y_train, y_train_min))))
print("Train set MAPE: " + str(metrics.mean_absolute_percentage_error(y_train, y_train_min)))

# Metrics of Training Set
Train set MSE: 11.6558110380783
Train set RMSE: 3.412604802893
```

Gambar 4. 27 Implementasi Model PR Orde 1 berdasarkan Data Latih


```

In [54]: # Metrics of Testing Set
print("Metrics of Testing Set")
y_hat_test = lin.predict(x_test)
print("Test set RMSE: ", np.sqrt(metrics.mean_squared_error(y_hat_test, y_test)))
print("Test set MAE: ", np.sqrt(metrics.mean_squared_error(y_hat_test, y_test)))
print("Test set MAPE: ", np.sqrt(metrics.mean_squared_error(y_hat_test, y_test)))

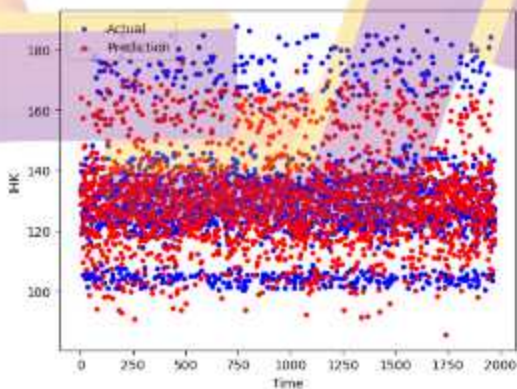
Metrics of Testing Set:
Test set RMSE: 316.093239524236
Test set MAE: 18.4542000209571
Test set MAPE: 8.46388713348884

```

Gambar 4. 28 Implementasi Model PR Orde 1 berdasarkan Data Uji

Tabel 4. 18 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 1

Tempat	Waktu	Data Sebenarnya	Data Hasil Prediksi
Jember	Januari 2014	164,8	157,0354
Jember	Februari 2014	168,53	162,6269
Jember	Maret 2014	172,92	162,8232
Jember	April 2014	171,31	164,7891
Jember	Mei 2014	167,62	165,3217
Jember	Juni 2014	168,5	162,1291
Jember	Juli 2014	177,12	160,6939
i	i	i	i
Surabaya	Desember 2020	104,61	120,844



Gambar 4. 29 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 1

b. Implementasi PR menggunakan Orde 2

Percobaan regresi polinomial selanjutnya dilakukan dengan Orde 2 (Gambar 4. 30). Sama seperti pada Orde 1, Orde 2 juga dilakukan evaluasi dengan MSE, RMSE, dan MAPE (Gambar 4. 31 dan Gambar 4. 32). Setelah proses evaluasi, dilanjutkan dengan visualisasi perbandingan antara data aktual dan data prediksi pada Tabel 4.19 dan Gambar 4.33.

```

In (34): # fitting polinomial regresi ke data aktual
fit2 = PolynomialRegression(degree = 2, fit_method =
'least_squares')
fit2.fit(x_train, y_train)

# hitung MSE
MSE = sum((y_train - fit2.predict(x_train))**2) / len(y_train)

```

Gambar 4. 30 Implementasi Model PR Orde 2

```

In (35): # hitung uji training set
print('MSE of training set:')
MSE_train = sum((y_train - fit2.predict(x_train))**2) / len(y_train)

print('MSE test set:')
MSE_test = sum((y_test - fit2.predict(x_test))**2) / len(y_test)

print('MSE test RMSE:')
RMSE_test = sqrt(MSE_test)

# hitung MSE, RMSE, MAPE
MSE = sum((y_train - fit2.predict(x_train))**2) / len(y_train)
RMSE = sqrt(MSE)
MAPE = sum(abs(y_train - fit2.predict(x_train)) / y_train) / len(y_train)

```

Gambar 4. 31 Implementasi Model PR Orde 2 berdasarkan Data Latih

```

In (37): # hitung uji testing set
print('MSE of testing set:')
MSE_test = sum((y_test - fit2.predict(x_test))**2) / len(y_test)

print('MSE test RMSE:')
RMSE_test = sqrt(MSE_test)

print('MSE test MAPE:')
MAPE_test = sum(abs(y_test - fit2.predict(x_test)) / y_test) / len(y_test)

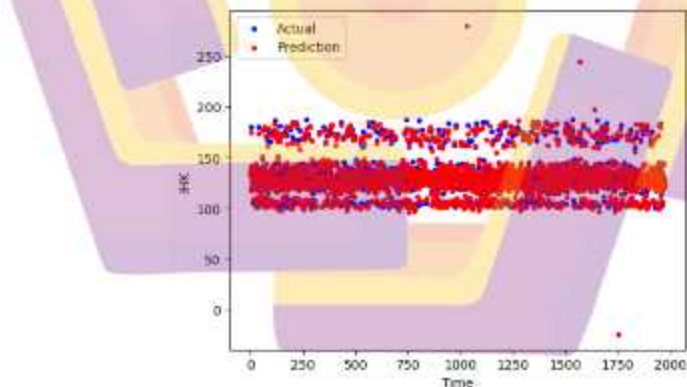
# hitung MSE, RMSE, MAPE
MSE = sum((y_train - fit2.predict(x_train))**2) / len(y_train)
RMSE = sqrt(MSE)
MAPE = sum(abs(y_train - fit2.predict(x_train)) / y_train) / len(y_train)

```

Gambar 4. 32 Implementasi Model PR Orde 2 berdasarkan Data Uji

Tabel 4. 19 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 2

Tempat	Waktu	Data Sebenarnya	Data Hasil Prediksi
Jember	Januari 2014	164,8	157,0345
Jember	Februari 2014	168,53	162,6269
Jember	Maret 2014	172,92	162,8232
Jember	April 2014	171,31	164,7891
Jember	Mei 2014	167,62	165,3217
Jember	Juni 2014	168,5	162,1291
Jember	Juli 2014	177,12	160,6939
⋮	⋮	⋮	⋮
Surabaya	Desember 2020	104,61	103,310



Gambar 4. 33 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 2

c. Implementasi PR menggunakan Orde 3

Percobaan terakhir dilakukan dengan Orde 3 (Gambar 4. 34). Hasil evaluasi regresi polinomial dengan orde ini ditampilkan pada Gambar 4. 35 dan Gambar 4. 36. Berdasarkan hasil visualisasi data dengan perbandingan data aktual dan hasil prediksi didapatkan plot yang hampir sejajar untuk semua data (Tabel 4. 20 dan Gambar 4. 37).

```

In [37]: # Menghitung koefisien Regresi orde 3 dan Akurasi
pdata = PolyaModel(akurasi=0.99, n = 3, degree =
k_order = polyn(3).coefficients)

# Hitung hasil
fdata = LinearRegression()
fdata.fit(xdata, ydata)

```

Gambar 4. 34 Implementasi Model PR Orde 3

```

In [38]: # Statistik of Training Set
print("Statistics of Training Set")
x_train_size = data.shape[0]
print("Train set Size: ", x_train_size, " y_train_size: ", y_train_size)
print("Train set MSE: ", sum((ydata - fdata.predict(xdata))**2) / x_train_size)
print("Train set RMSE: ", sqrt(sum((ydata - fdata.predict(xdata))**2) / x_train_size))

# Statistik of Testing Set
train_set_size = 0.7 * data.shape[0]
test_set_size = data.shape[0] - train_set_size
test_set_size = 0.3 * data.shape[0]

```

Gambar 4. 35 Implementasi Model PR Orde 3 berdasarkan Data Latih

```

In [39]: # Statistik of Testing Set
print("Statistics of Testing Set")
x_test_size = data.shape[0] * (1 - 0.7)
print("Test set Size: ", x_test_size, " y_test_size: ", y_test_size)
print("Test set MSE: ", sum((ydata - fdata.predict(x_test))**2) / x_test_size)
print("Test set RMSE: ", sqrt(sum((ydata - fdata.predict(x_test))**2) / x_test_size))
print("Test set MAE: ", sum(abs(ydata - fdata.predict(x_test))) / x_test_size)

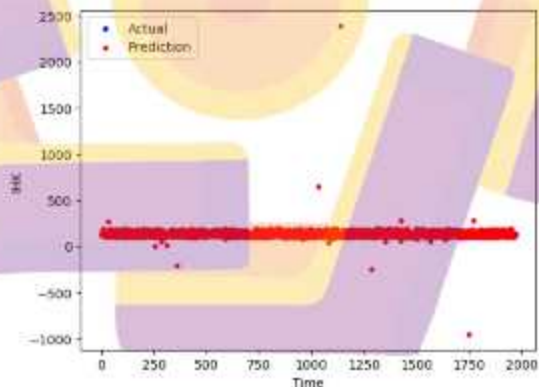
# Statistik of Testing Data
print("Statistics of Testing Data")
test_set_size = 0.3 * data.shape[0]
test_set_size = 0.3 * data.shape[0]
test_set_size = 0.3 * data.shape[0]
test_set_size = 0.3 * data.shape[0]

```

Gambar 4. 36 Implementasi Model PR Orde 3 berdasarkan Data Uji

Tabel 4. 20 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 3

Tempat	Waktu	Data Sebenarnya	Data Hasil Prediksi
Jember	Januari 2014	164,8	157,0345
Jember	Februari 2014	168,53	162,6269
Jember	Maret 2014	172,92	162,8232
Jember	April 2014	171,31	164,7891
Jember	Mei 2014	167,62	165,3217
Jember	Juni 2014	168,5	162,1291
Jember	Juli 2014	177,12	160,6939
⋮	⋮	⋮	⋮
Surabaya	Desember 2020	104,61	104,67



Gambar 4. 37 Perbandingan Data Sebenarnya dan Data Hasil Prediksi menggunakan PR Orde 3

Berdasarkan hasil implementasi model PR untuk prediksi IHK dengan Orde 1, 2, dan 3 dapat disimpulkan hasil evaluasi seperti pada Tabel 4. 21.

Tabel 4. 21 Hasil Evaluasi Model PR menggunakan MSE, RMSE, dan MAPE

Kriteria	Orde 1		Orde 2		Orde 3	
	Train	Test	Train	Test	Train	Test
MSE	124.454	118.690	17.548	50.801	2.136	3530.197
RMSE	11.155	10.894	4.189	7.127	1.461	59.415
MAPE	0.064	0.063	0.021	0.024	0.003	0.034
Compile Time (Sec)	0.044	0.004	1.817	0.041	144.591	0.220

Pada percobaan pertama menggunakan Orde 1, nilai MSE untuk data latih dan data uji mencapai 124.454 dan 118.690. Hasil RMSE untuk data latih dan data uji berturut-turut menghasilkan nilai 11.155 dan 10.894. Hasil evaluasi dari kedua metode ini cenderung sangat tinggi dan jauh dari 0. Sedangkan pada evaluasi menggunakan MAPE menghasilkan nilai 0,064 (6,4%) dan 0,063 (6,3%).

Percobaan kedua dengan Orde 2, untuk hasil evaluasi MSE dan RMSE memiliki performa yang jauh lebih baik daripada Orde 1. Nilai MSE untuk data latih dan data uji mencapai 17.548 dan 50.801. Nilai RMSE mencapai 4.189 dan 7.127. Namun meskipun sudah jauh lebih rendah daripada percobaan pada Orde 1, nilai yang terbentuk pada Orde 2 juga masih jauh dari 0. Nilai MAPE yang terbentuk pada orde ini pun lebih baik daripada Orde 1, dimana percobaan untuk data latih menghasilkan nilai 0,021 atau setara 2,1% dan percobaan pada data latih menghasilkan nilai 0,024 atau setara 2,4%.

Hasil terbaik untuk implementasi model jatuh pada Orde 3, nilai MAPE untuk percobaan dengan data latih mencapai 0,3% dan data ujinya mencapai 3,4%. Jika dilihat dari visualisasi perbandingan hasil prediksi dan data sebenarnya pada

Gambar 4. 37, grafik pun menunjukkan garis lurus yang hampir sempurna. Sehingga dapat disimpulkan proses implementasi model menggunakan regresi polinomial memiliki performa terbaik pada Orde 3.

Penilaian performa selanjutnya ditunjukkan melalui waktu tunggu dalam proses implementasi model. Pada Orde 1, waktu tunggu mencapai 0,044 sekon untuk perobaan menggunakan data latih dan percobaan menggunakan data uji. Pada Orde 2, waktu tunggu mencapai 1,187 dan 0,041 masing-masing untuk data latih dan data uji. Sedangkan pada Orde 3, waktu tunggu untuk percobaan dengan data latih adalah 144,591 sekon dan 0,220 sekon untuk percobaan dengan data uji.

4.2.4. Perbandingan Prediksi IHK menggunakan SVR dan PR

Pada sub bab ini akan dijelaskan kesimpulan hasil implementasi model menggunakan SVR dan PR untuk prediksi Nilai IHK. Baik pada model SVR maupun PR memiliki tingkat kesalahan yang rendah, yaitu di bawah 10%. Pada algoritma prediksi, jika tingkat kesalahan yang dihasilkan <10%, maka algoritma tersebut memiliki kehandalan yang baik. Dapat disimpulkan pula bahwa algoritma SVR maupun PR berhasil digunakan dengan baik untuk proses prediksi nilai IHK.

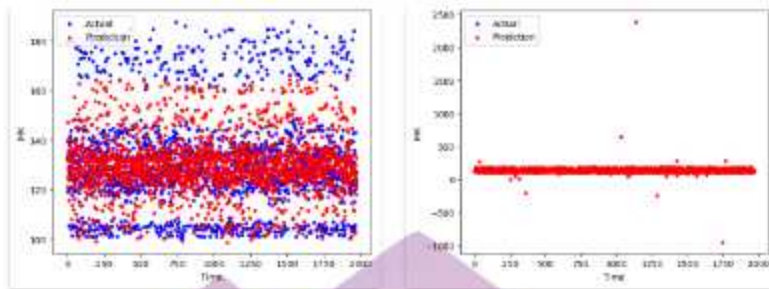
Pada implementasi model SVR, digunakan satu kernel yaitu RBF. Hasil terbaik ditunjukkan pada perbandingan data latih dan data uji 90%:10%. Nilai MAPE yang dihasilkan adalah 5,9% untuk percobaan data latih dan 6% untuk data uji. Sementara pada implementasi model PR dilakukan tiga kali percobaan yang menghasilkan performa paling baik pada Orde 3 dengan nilai MAPE 0,3% pada percobaan data latih dan 3,4% pada percobaan data uji. Hasil perbandingan kedua

model dapat dilihat pada Tabel 4. 22. Berdasarkan hasil yang terlihat di tabel, dapat disimpulkan bahwa model regresi polinomial atau PR memiliki performa yang lebih baik daripada SVR dengan nilai MAPE yang lebih rendah atau mendekati nilai 0.

Tabel 4. 22 Perbandingan Hasil SVR dan PR Orde 3

Kriteria	SVR		PR Orde 3	
	Train	Test	Train	Test
MSE	137.025	139.775	2.136	3530.197
RMSE	11.705	11.822	1.461	59.415
MAPE	0.059	0.060	0.003	0.034
Compile Time	740,403	5,229	144,591	0,220

Perbandingan selanjutnya dilakukan untuk membandingkan algoritma SVR dan PR dalam kecepatan proses implementasi modelnya. SVR menghabiskan 740,403 sekon dan 5,229 sekon untuk percobaan data latih dan data uji. Hasil *compile time* untuk PR juga lebih baik, hanya membutuhkan waktu 144,591 sekon untuk percobaan data latih dan 0,220 sekon untuk data uji. Dapat disimpulkan secara keseluruhan bahwa model PR dengan Orde 3 mampu mengungguli model SVR dalam proses prediksi nilai IHK.



Gambar 4. 38 Perbandingan Plot Data Sebenarnya dan Data Hasil Prediksi SVR dan PR Orde 3

Jika dilihat dari perbandingan visualisasi hasil prediksi menggunakan SVR dan PR Orde 3 dapat dilihat *gap* yang cukup besar (Gambar 4. 38). Pada kasus penelitian sebelumnya yang melakukan implementasi SVR untuk regresi, tingkat kesalahan yang dihasilkan cenderung rendah atau memiliki performa yang memuaskan (Li *et al.*, 2022). Namun dalam prediksi nilai IHK ternyata tingkat kesalahan yang dihasilkan masih lebih tinggi daripada PR Orde 3. Hal ini disebabkan karena dalam implementasi SVR diperlukan variabel yang memiliki korelasi tinggi antar satu sama lain. Sementara dalam penelitian ini, total 34 komoditas dengan nilai IHK memiliki korelasi dalam rentang sangat rendah, rendah, dan sedang.

Dalam implementasi PR, beberapa penelitian yang dilakukan sebelumnya pun sudah menggunakan variabel dengan korelasi yang tinggi. Namun hasil yang ditunjukkan tidak selalu memuaskan. Pada penelitian Erfiana, tingkat kesalahan yang ditunjukkan mencapai 36,6% (Erfiana, 2015). Sedangkan dalam penelitian yang lain, PR dibandingkan dengan Regresi Linier dengan hasil Regresi Linier yang

lebih unggul (Eka *et al.*, 2021). Dalam penelitian ini korelasi yang terbentuk antar variabel cenderung jauh dari 1 dan -1, yang artinya cenderung rendah. Namun ternyata hasil prediksi menggunakan PR Orde 3 memiliki tingkat kesalahan yang rendah. Seperti yang telah dipaparkan sebelumnya, bahwa kekurangan dari PR adalah belum bisa menjamin performa hasil prediksi dengan baik. Ketika menggunakan orde yang tinggi, maka galat akan menjadi kecil atau bahkan hilang sama sekali. Namun hal ini tidak bisa mencerminkan keandalan model yang digunakan. Jika derajat yang dipilih rendah, maka galatnya akan besar. Jika derajat yang dipilih tinggi, maka galat akan menjadi kecil atau bahkan tidak memiliki galat.

Tujuan utama yang ingin dicapai dalam PR adalah membangun model yang sederhana untuk melakukan proses prediksi. Semakin tinggi derajat polinomial yang digunakan, maka semakin kompleks pula model yang didapat. Namun model yang kompleks juga tidak mencerminkan keandalan suatu model prediksi. Karena nilai galat yang tercipta mungkin hanya baik pada data yang sedang digunakan, dan memungkinkan bahwa tidak akan sama hasilnya jika digunakan untuk data yang lain.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian yang sudah dibahas pada bab sebelumnya, dapat disimpulkan beberapa hal sebagai berikut:

1. Analisa korelasi menggunakan Pearson Product Moment. Hasil Analisa korelasi menunjukkan adanya korelasi negatif sempurna antara variabel x (harga komoditas makanan) terhadap variabel y (nilai IHK). Hal ini menunjukkan hubungan yang berbalik, yaitu jika variabel x naik, maka variabel y turun. Artinya, jika harga komoditas naik, maka nilai IHK turun. Meskipun mayoritas komoditas memiliki korelasi negatif dengan nilai IHK, namun ada beberapa komoditas yang memiliki korelasi positif. Cabe Rawit, Kacang Kedelai Eks Impor, Minyak Curah, Minyak Goreng Bimoli Botol/Kemasan 2-ltr, dan Susu Bubuk Merk Indomilk (Instant) merupakan komoditas yang memiliki korelasi positif. Berdasarkan hasil penelitian ini, komoditas yang memiliki korelasi positif artinya memiliki hubungan yang lurus. Semakin naik variabel x nya, maka semakin naik pula variabel y nya;
2. Prediksi nilai IHK dilakukan dengan dua model, yaitu SVR dan PR. Hasil evaluasi model menggunakan MSE dan RMSE kurang memuaskan, dengan nilai yang sangat jauh dari 0. Sementara hasil evaluasi menggunakan MAPE menunjukkan nilai yang cukup memuaskan. Nilai MAPE yang dihasilkan model SVR adalah 5,9% untuk percobaan data latih dan 6% untuk data uji.

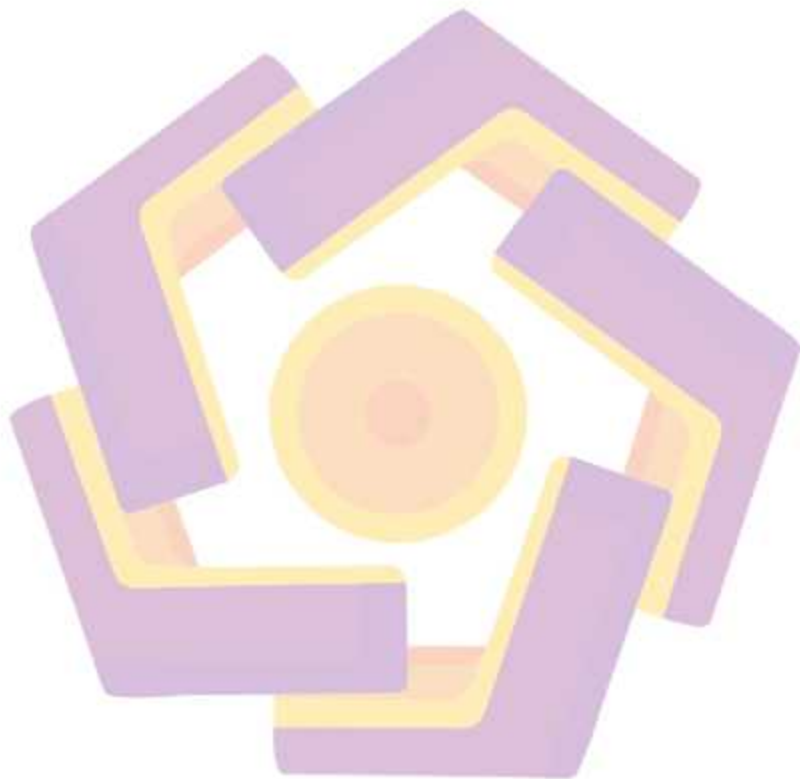
Sedangkan model PR menghasilkan nilai MAPE 0,3% pada percobaan data latih dan 3,4% pada percobaan data uji (Orde 3). Berdasarkan hasil evaluasi menggunakan MAPE dapat disimpulkan bahwa model terbaik untuk prediksi nilai IHK adalah Polynomial Regression (PR) dengan Orde 3. Sedangkan berdasarkan *compile time* nya, SVR membutuhkan 740,403 sekon dan 5,229 sekon untuk percobaan data latih dan data uji. Sedangkan PR hanya membutuhkan waktu 144,591 untuk percobaan data latih dan 0,220 untuk data uji. Dapat disimpulkan bahwa prediksi menggunakan SVR dan PR berhasil dilakukan untuk melakukan prediksi IHK, dengan PR Orde 3 sebagai model yang lebih unggul.

5.2. Saran

Beberapa saran yang dapat dilakukan pada penelitian selanjutnya:

1. Implementasi model dilakukan dengan memanfaatkan data harga komoditas makanan di Provinsi Jawa Timur. Pada penelitian selanjutnya dapat dilakukan dengan memanfaatkan data dari seluruh Indonesia untuk mendapatkan hasil yang lebih baik;
2. Proses prediksi bisa dilakukan dengan lebih banyak data harga komoditas, tidak hanya komoditas makanan saja agar dapat menarik kesimpulan yang lebih besar;
3. Menggunakan kernel lain selain RBF, seperti Kernel Linear dan Kernel Polinomial;

4. Metode evaluasi menggunakan MSE dan RMSE memiliki tingkat kesalahan yang sangat tinggi dan jauh dari 0. Pada penelitian selanjutnya dapat menggunakan metode evaluasi yang lain agar hasil prediksi menjadi lebih akurat.



DAFTAR PUSTAKA

PUSTAKA BUKU

Bernhard Scholkopf, A.J.S. (2018) *Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press. Available at: https://www.google.co.id/books/edition/Learning_with_Kernels/7r34DwAAQBAJ?hl=id&gbpv=0.

Sugiyono (2012) *Metode Penelitian Kuantitatif dan R&D*. Bandung: Alfabeta.

PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

Ananda, I.A.R., Tarno, T. and Sudarno, S. (2020) 'Peramalan Data Indeks Harga Konsumen Kota Purwokerto Menggunakan Model Fungsi Transfer Multi Input', *Jurnal Gaussian*, 9(4), pp. 515–524. doi:10.14710/j.gauss.v9i4.29406.

Aryani, L. et al. (2020) 'Prediksi jumlah siswa baru dengan menggunakan metode exponential smoothing (studi kasus: SMK Ethika Palembang)', *Bina Darma Conference on Computer ...*, pp. 237–244. Available at: <https://conference.binadarma.ac.id/index.php/BDCCS/article/view/1615%0Ahttps://conference.binadarma.ac.id/index.php/BDCCS/article/download/1615/693>.

Asyiva, A. (2019) 'Prediksi laju inflasi indonesia menggunakan metode support vector regression dengan kernel radial basis function'.

Ayudhiah, M.P., Bahri, S. and Fitriyani, N. (2020) 'Peramalan Indeks Harga Konsumen Kota Mataram Menggunakan Vector Autoregressive Integrated Moving Average', *Eigen Mathematics Journal*, 1(2), p. 1. doi:10.29303/emj.v1i2.61.

Azhar, S. (2007) 'Improving collaboration between researchers and practitioners in construction research projects using action research technique', *Associated Schools of Construction, International ...* [Preprint], (August). Available at: <http://ascpro0.ascweb.org/archives/2007/CERT81002007.pdf>.

Chang, P.-C., Wang, Y.-W. & Liu, C.-H. (2007) *The Development of a Weighted Evolving Fuzzy Neural Network for PCB Sales Forecasting*. 32 (Expert. Elsevier.

Dewi, C., Kartikasari, D.P. and Mursityo, Y.T. (2014) 'PREDIKSI CUACA PADA DATA TIME SERIES MENGGUNAKAN ADAPTIVE NEURO FUZZY

INFERENCE SYSTEM (ANFIS)', 1(1), pp. 18–24.

Dimashanti, A.R. and Sugiman (2021) 'Peramalan Indeks Harga Konsumen Kota Semarang Menggunakan SARIMA Berbantuan Software Minitab', *Prisma*, 4, pp. 565–576. Available at: <https://journal.unnes.ac.id/sju/index.php/prisma/>.

Eka, A. *et al.* (2021) 'Prediksi Produksi Daging Sapi Nasional dengan Metode Regresi Linier dan Regresi Polinomial Pendahuluan Regresi Linier Regresi Polinomial', 20, pp. 209–215.

Hali Mukron, M. *et al.* (2021) 'Peramalan Indeks Harga Konsumen Indonesia Menggunakan Autoregressive Integrated Moving Average', *Jurnal Statistika Industri dan Komputasi*, 6(1), pp. 20–25.

Jassim, M.S., Coskuner, G. and Zontul, M. (2022) 'Comparative performance analysis of support vector regression and artificial neural network for prediction of municipal solid waste generation', *Waste Management and Research*, 40(2), pp. 195–204. doi:10.1177/0734242X211008526.

Li, Q. *et al.* (2022) 'State of health estimation of lithium-ion battery based on improved ant lion optimization and support vector regression', *Journal of Energy Storage*, 50. doi:10.1016/j.est.2022.104215.

Maricar, A.M. (2019) 'Analisa Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ', *Jurnal Sistem dan Informatika (JSI)*, 13(2), pp. 36–45. Available at: <https://www.jsi.stikom-bali.ac.id/index.php/jsi/article/view/193>.

Maulana, N.D., Setiawan, B.D. and Dewi, C. (2019) 'Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus : Harum Bakery)', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(3), pp. 2986–2995.

Nabillah, I. and Ranggadara, I. (2020) 'Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut', *JOINS (Journal of Information System)*, 5(2), pp. 250–255. doi:10.33633/joins.v5i2.3900.

Rohmah, F.M. *et al.* (2019) 'Meramal Indeks Harga Konsumen Kabupaten di Jawa Timur dengan Metode Support Vector Regression Data Mining', *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, (x), pp. 30–36.

Sande, S. and Privalsky, M.L. (1996) 'Identification of TRACs (T3 receptor-associating cofactors), a family of cofactors that associate with, and modulate the activity of, nuclear hormone receptors', *Molecular Endocrinology*,

10(7), pp. 813–825. doi:10.1210/me.10.7.813.

Wahyu Widayati, C.S. (2013) 'Komparasi Beberapa Metode Estimasi Kesalahan Pengukuran', *Jurnal Penelitian dan Evaluasi Pendidikan*, 13(2), pp. 182–197. doi:10.21831/pep.v13i2.1409.

Zahara, S. and Sugianto (2021) 'Peramalan Data Indeks Harga Konsumen Berbasis Time Series Multivariate Menggunakan Deep Learning', *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(1), pp. 24–30. doi:10.29207/resti.v5i1.2562.

Zahara, S. and Sugianto, S. (2021) 'Prediksi Indeks Harga Konsumen Komoditas Makanan Berbasis Cloud Computing Menggunakan Multilayer Perceptron', *JOINTECS (Journal of Information Technology and Computer Science)*, 6(1), p. 21. doi:10.31328/jointecs.v6i1.1702.

PUSTAKA LAPORAN PENELITIAN

Erfiana, N. (2015) 'EKSPLORASI MINYAK DAN GAS BUMI (Studi Kasus : Blok Banyu Urip Kab , Bojonegoro , Jawa Timur)', *Tesis Jurusan Teknik Geomatika* [Preprint].

Hasan, N.F. (2020) *Tesis penerapan jaringan syaraf tiruan dengan algoritma*.

Susianto, L.J. (2016) 'Perbandingan Model Regresi Polinomial dan Model Regresi Kernel Nadaraya-Watson: Studi Kasus Harga Emas di Indonesia', *Skripsi* [Preprint].

PUSTAKA ELEKTRONIK

Afifah, L. (2022) *Apa Itu Bias dan Variance di Machine Learning?*, *ilmudataPy*. Available at: <https://ilmudatapy.com/apa-itu-bias-dan-variance-di-machine-learning/> (Accessed: 11 January 2023).

Statistik, B.P. (2023) *Indeks Harga Konsumen (IHK)*. Available at: <https://www.cnnindonesia.com/ekonomi/20220210120551-537-757460/mengenal-indeks-harga-konsumen-dan-kegunaannya> (Accessed: 11 January 2023).

Timur, B.P.S.P.J. (2020) *Indeks Harga Konsumen 8 Kota di Provinsi Jawa Timur 2020*. Available at: <https://jatim.bps.go.id/publication/2021/04/23/4932c8cfa622e3c2618e04d8/indeks-harga-konsumen-8-kota-di-provinsi-jawa-timur-2020.html>.

