

LAPORAN PENELITIAN KELOMPOK
Studi Komprehensif tentang Fluktuasi Harga Cabai Rawit Merah di Pasar
Tradisional dan Modern Jakarta



PENELITI

RAVANSA RAHMAN SANTOSA	312110103
AJIE RAFLI PAMUNGKAS	312110429
MUHAMMAD KHRISNA FAISAL ZUHRI	312110177
ANINDIA SASI KIRANA	312110268
MUFIDA NURIYANA	312110172

UNIVERSITAS PELITA BANGSA FAKULTAS TEKNIK PRODI
TEKNIK INFORMATIKA

2021

KATA PENGANTAR

Puji syukur kita panjatkan kehadiran Allah Swt. yang telah memberikan rahmat dan hidayah-Nya sehingga saya dapat menyelesaikan laporan penelitian yang berjudul “Studi Komprehensif tentang Fluktuasi Harga Cabai Rawit Merah di Pasar Tradisional dan Modern Jakarta” ini tepat pada waktunya.

Adapun tujuan dari penulisan dari laporan ini adalah untuk memenuhi tugas pada mata kuliah Big Data. Selain itu, laporan ini juga bertujuan untuk Semoga hasil studi ini dapat memberikan wawasan yang berharga bagi para pelaku industri, pengambil kebijakan, dan masyarakat umum. Terlebih dahulu, kami mengucapkan terima kasih kepada Bapak Dr. Ir. Ananto Tri Sasongko, M.Sc. selaku Dosen Big Data yang telah memberikan tugas ini sehingga dapat menambah pengetahuan dan wawasan sesuai dengan bidang studi yang kami tekuni ini.

Kemudian, saya menyadari bahwa tugas yang saya tulis ini masih jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun kami butuhkan demi kesempurnaan laporan ini.

Bekasi, Januari 2024

Tim Peneliti

DAFTAR ISI

Halaman Judul	i
Kata Pengantar	ii
Daftar Isi	iii
Abstrak	iv
Bab I Pendahuluan	
1.1 Latar Belakang	4
Bab II Tinjauan Pustaka	
2.1 Big Data.....	5
2.2 Mengambil Data	6
2.3 Pembersih Data	6
2.4 Modeling	7
2.5 Interpretasi/Evaluasi	7
2.6 Pyspark	9
2.7 Regresi	9
2.8 Linear Regression	9
BAB III METODE PENELITIAN	
3.1 Rancangan Penelitian	10
3.2 Mengambil Data	10
3.3 Import Library	10
3.4 Load Dataset	10
3.5 Analisis Data	11
3.6 Split Dataset	11
3.7 Membangun Model Linear Regression	11
3.8 Evaluasi	12
BAB IV HASIL DAN ANALISA	
4.1 Pengumpulan Data	13
4.2 Analisis Data	14
4.3 Splitting Data	19
4.4 Membangun Model Linear Regression	19
4.5 Evaluasi	20
BAB VI KESIMPULAN	
DAFTAR PUSTAKA	

ABSTRAK

Cabai rawit merah memiliki peran penting dalam industri pertanian di Indonesia, memengaruhi rasa dan aroma masakan lokal. Fluktuasi harga cabai rawit merah menjadi perhatian utama untuk menjaga stabilitas ekonomi dan ketahanan pangan. Penelitian ini bertujuan untuk memahami fluktuasi harga cabai rawit merah di pasar tradisional dan modern di Jakarta dengan menggunakan pendekatan analisis data terdistribusi menggunakan PySpark. Data harga cabai rawit merah dari tahun 2020 hingga 2023 diambil dari Info Pangan Jakarta. Analisis data melibatkan pemahaman pola, tren, dan evaluasi statistik untuk empat pasar yang berbeda: Pasar Induk Kramat Jati, Pasar Grogol, Pasar Minggu, dan Pasar Koja Baru. Hasil analisis deskriptif menunjukkan variasi harga yang signifikan di Pasar Koja Baru dan konsistensi harga yang lebih tinggi di Pasar Induk Kramat Jati. Visualisasi time series menunjukkan peningkatan rata-rata nilai transaksi pasar tradisional di Jakarta selama tiga tahun terakhir.

Pada tahap pemodelan, digunakan metode Linear Regression untuk memprediksi harga cabai rawit merah. Evaluasi model menunjukkan tingkat akurasi yang tinggi (R^2 : 95%) dan tingkat kesalahan prediksi yang rendah (MSE dan MAE yang rendah). Penelitian ini memberikan wawasan mendalam mengenai fluktuasi harga cabai rawit merah, potensi faktor-faktor pengaruh, dan strategi manajemen risiko. Hasilnya dapat bermanfaat bagi petani, pedagang, serta kebijakan ekonomi dan pertanian di tingkat lokal maupun nasional.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Cabai rawit merah merupakan salah satu komoditas penting dalam industri pertanian di Indonesia, memiliki peran signifikan dalam menciptakan rasa dan aroma khas pada berbagai masakan lokal. Ketersediaan dan fluktuasi harga cabai rawit merah menjadi perhatian utama para pelaku usaha, konsumen, serta pemerintah yang ingin menjaga stabilitas ekonomi dan ketahanan pangan. Penelitian ini bertujuan untuk melakukan studi komprehensif terkait fluktuasi harga cabai rawit merah di pasar tradisional dan modern di Jakarta. Pasar tradisional dan modern masing-masing memiliki karakteristik dan dinamika tersendiri yang memengaruhi perubahan harga komoditas pertanian. Dengan memahami perbedaan dan kesamaan dalam fluktuasi harga di kedua jenis pasar tersebut, diharapkan penelitian ini dapat memberikan wawasan yang lebih mendalam mengenai faktor-faktor yang mempengaruhi dinamika harga cabai rawit merah di Ibu Kota. Keberhasilan penelitian ini dapat memberikan manfaat luas, tidak hanya bagi para petani dan pedagang cabai rawit merah, tetapi juga untuk kebijakan ekonomi dan pertanian di tingkat lokal maupun nasional. Dengan memahami pola fluktuasi harga, pemangku kepentingan dapat mengembangkan strategi yang lebih efektif dalam manajemen risiko, penentuan harga yang adil, serta perencanaan bisnis yang berkelanjutan dalam rantai pasok cabai rawit merah.

BAB II

TINJAUAN PUSTAKA

2.1 Big Data

Big data adalah istilah yang merujuk pada volume data yang sangat besar, kompleksitas tinggi, dan beragam jenisnya yang sulit atau bahkan tidak dapat diolah menggunakan metode tradisional pemrosesan data. Konsep ini mencakup karakteristik utama yang dikenal sebagai "3V": Volume, Velocity, dan Variety.

1. Volume (Besar): Merupakan karakteristik yang mengacu pada jumlah besar data yang terlibat dalam big data. Ini dapat mencakup data dari berbagai sumber seperti sensor, perangkat IoT (Internet of Things), log transaksi, media sosial, dan banyak lagi. Jumlah data yang besar ini seringkali melebihi kapasitas penyimpanan dan kemampuan pemrosesan data tradisional.
2. Velocity (Cepat): Big data sering kali mengalir dengan kecepatan tinggi. Data dapat dihasilkan atau diperbarui secara real-time atau mendekati waktu nyata, seperti data transaksi keuangan, data sensor, dan aliran informasi media sosial. Kemampuan untuk memproses dan merespons data dengan cepat menjadi kritis untuk mendukung pengambilan keputusan yang real-time.
3. Variety (Beragam): Karakteristik ini menunjukkan bahwa data big data dapat berasal dari berbagai sumber dan dalam berbagai format. Data bisa terstruktur (seperti data dalam database relasional) atau tidak terstruktur (seperti teks, gambar, audio, dan video). Keanekaragaman jenis data ini menantang karena memerlukan alat dan teknik pemrosesan data yang mampu mengatasi berbagai format dan struktur.

Selain 3V, terdapat dimensi tambahan dalam konsep big data:

1. Veracity (Keandalan): Merujuk pada kualitas data. Dalam big data, sumber data dapat tidak pasti atau tidak selalu akurat. Oleh karena itu, penting untuk menilai dan memastikan keandalan data yang digunakan untuk analisis.
2. Variability (Variabilitas): Menunjukkan fluktuasi dalam data yang dapat berubah seiring waktu. Data big data mungkin mengalami perubahan dalam pola atau karakteristiknya, dan pemrosesan data harus dapat menangani variabilitas ini.

3. Value (Nilai): Akhirnya, nilai adalah inti dari konsep big data. Tujuan dari mengelola dan menganalisis big data adalah untuk mendapatkan wawasan yang berharga dan informasi yang dapat mendukung pengambilan keputusan yang lebih baik.

2.2 Mengambil Data

Pertanian memiliki peran sentral dalam perekonomian Indonesia, dengan berbagai jenis komoditas yang menjadi tulang punggung sektor ini. Studi sebelumnya tentang dinamika pasar pertanian, terutama pada komoditas sayuran dan rempah-rempah, memberikan landasan untuk memahami peran dan tantangan pasar pertanian di tingkat nasional.

2.3 Pembersihan Data

Pembersihan data (data cleansing atau data cleaning) merujuk pada serangkaian proses untuk mengidentifikasi, memperbaiki, dan menghapus kesalahan atau inkonsistensi dalam satu set data. Tujuan utama dari pembersihan data adalah untuk meningkatkan kualitas data, membuat data lebih akurat, lengkap, dan konsisten. Proses ini menjadi sangat penting karena data yang buruk atau tidak terawat dapat menghasilkan kesalahan dalam analisis dan pengambilan keputusan.

berikut adalah lima langkah utama dalam pembersihan data:

1. Deteksi Kesalahan: Identifikasi dan analisis kesalahan dalam data, seperti nilai yang hilang, duplikat, atau inkonsistensi.
2. Penanganan Duplikat: Menangani data duplikat dengan menggabungkan atau menghapus salinan yang tidak diperlukan.
3. Penanganan Nilai yang Hilang: Mengidentifikasi dan mengatasi nilai yang hilang atau kosong dalam data.
4. Validasi Data: Memverifikasi keakuratan data dengan memeriksa apakah nilai-nilai berada dalam rentang yang valid atau sesuai dengan aturan bisnis yang berlaku.
5. Normalisasi Data: Menyederhanakan data dengan mengonversi format atau satuan yang berbeda untuk memastikan konsistensi dan kemudahan perbandingan.

2.4 Modeling

Modeling (modeling data atau modeling informasi) adalah proses pembuatan representasi struktural dari suatu konsep atau sistem dengan menggunakan simbol, aturan, dan notasi tertentu. Dalam berbagai konteks, istilah "modeling" dapat merujuk pada berbagai jenis model, seperti:

1. Model Matematika: Representasi matematis dari suatu fenomena atau sistem dengan menggunakan persamaan atau fungsi matematika.
2. Model Komputer: Representasi digital dari konsep atau sistem yang dapat dijalankan pada komputer, umumnya digunakan untuk simulasi atau pengujian.
3. Model Data: Representasi struktur data yang digunakan dalam basis data untuk menyimpan dan mengakses informasi.
4. Model Bisnis: Representasi struktur atau proses bisnis suatu organisasi, seringkali digunakan untuk perencanaan dan analisis proses bisnis.
5. Model Prediktif (dalam ilmu data dan pembelajaran mesin): Pembuatan model yang dapat membuat prediksi atau klasifikasi berdasarkan pola yang ditemukan dalam data.

Modeling dapat digunakan dalam berbagai disiplin ilmu dan industri untuk memahami, merancang, dan mengelola suatu sistem atau konsep. Dalam konteks ilmu data dan pembelajaran mesin, modeling juga merujuk pada pembuatan model prediktif berdasarkan data, di mana algoritma belajar mesin digunakan untuk mengekstraksi pola dari data dan membuat prediksi atau klasifikasi

2.5 Interpretasi/Evaluasi

Interpretasi atau evaluasi adalah langkah kritis dalam analisis data, di mana hasil analisis atau model diinterpretasikan dan dinilai untuk mendapatkan pemahaman yang lebih dalam atau mengevaluasi kualitasnya. Dalam berbagai konteks, interpretasi dan evaluasi dapat memiliki arti yang sedikit berbeda:

1. Analisis Data dan Statistik:
 - Interpretasi Data: Mengartikan temuan atau pola yang dihasilkan dari analisis data.

Ini melibatkan pemahaman makna dari hasil statistik atau visualisasi data.

- Evaluasi Statistik: Menilai keandalan dan relevansi analisis statistik. Evaluasi ini melibatkan pertimbangan terhadap metode statistik yang digunakan, ukuran keandalan (seperti nilai p-nilai), dan interpretasi kesalahan yang mungkin terjadi.
2. Pembelajaran Mesin:
- Interpretasi Model: Mengartikan bagaimana model pembelajaran mesin membuat prediksi atau klasifikasi. Ini melibatkan analisis bobot fitur, pentingnya variabel, dan bagaimana model membuat keputusan.
 - Evaluasi Model: Mengukur kinerja model berdasarkan berbagai metrik evaluasi, seperti akurasi, presisi, recall, dan area di bawah kurva ROC. Evaluasi ini membantu menentukan seberapa baik model dapat digeneralisasi ke data baru.
3. Evaluasi Proyek atau Inisiatif:
- Interpretasi Hasil Proyek: Mengartikan dampak atau hasil dari suatu proyek atau inisiatif. Ini melibatkan penafsiran terhadap pencapaian tujuan dan dampaknya pada pemangku kepentingan.
 - Evaluasi Keseluruhan Proyek: Menilai efektivitas, efisiensi, dan keberlanjutan proyek. Evaluasi ini mencakup penilaian terhadap penggunaan sumber daya, pencapaian tujuan, dan dampak jangka panjang.
4. Penelitian Ilmiah:
- Interpretasi Temuan Penelitian: Mengartikan hasil eksperimen atau temuan penelitian dan menyusun narasi yang menggambarkan makna dari temuan tersebut.
 - Evaluasi Metodologi: Menilai kualitas metodologi penelitian, keabsahan temuan, dan relevansinya terhadap literatur ilmiah yang ada.
5. Evaluasi Kinerja Bisnis:
- Interpretasi Kinerja: Mengartikan hasil kinerja bisnis, seperti penjualan, keuntungan, atau kepuasan pelanggan, untuk mendapatkan wawasan yang dapat digunakan dalam pengambilan keputusan.
 - Evaluasi Strategi Bisnis: Menilai efektivitas strategi bisnis dan menentukan apakah tujuan bisnis telah tercapai.

Interpretasi dan evaluasi membantu dalam menyusun pemahaman yang mendalam tentang data, model, atau hasil proyek, dan memastikan bahwa keputusan yang diambil didasarkan pada pemahaman yang solid dan relevan.

2.6 Pyspark

PySpark adalah antarmuka pemrograman Python untuk Apache Spark, suatu platform komputasi distribusi open-source yang dirancang untuk pemrosesan big data. Apache Spark menyediakan eksekusi tugas yang cepat dan umum untuk analisis data di dalam cluster komputasi. Dengan menggunakan PySpark, pengembang dapat memanfaatkan kekuatan Apache Spark menggunakan Python, yang membuatnya lebih mudah untuk digunakan dan diadopsi oleh para pengembang Python. Penggunaan PySpark membuka akses ke kekuatan dan kemampuan Apache Spark dalam pemrosesan data besar dan kompleks, sambil memanfaatkan fleksibilitas dan kenyamanan pemrograman Python.

2.7 Regresi

Penelitian sebelumnya telah mengidentifikasi faktor-faktor ekonomi dan sosial yang mempengaruhi fluktuasi harga pada pasar tradisional. Kajian ini akan membantu dalam mengidentifikasi variabel-variabel kunci yang perlu diperhatikan dalam konteks harga cabai rawit merah di pasar tradisional Jakarta.

2.8 Linear Regression

Studi pustaka tentang manajemen risiko dalam konteks harga komoditas pertanian akan memberikan wawasan terhadap strategi-strategi yang telah digunakan dalam mengelola fluktuasi harga, khususnya pada komoditas cabai rawit merah.

BAB III

METODE PENELITIAN

3.1 Rancangan Penelitian

Penelitian ini mengadopsi pendekatan analisis data terdistribusi menggunakan PySpark untuk memperoleh pemahaman tentang fluktuasi harga cabai rawit merah di pasar tradisional dan modern di Jakarta.

3.2 Mengambil Data

Kumpulan data harga cabai dari tahun 2020 hingga 2023 dengan periode harian diambil dari situs web Info Pangan Jakarta.

3.3 Import Library

Di sini mengimport library yang diperlukan, seperti yang tercantum di bawah ini:

```
from pyspark.sql import SparkSession
from pyspark.sql.types import StructType, StructField, DateType, FloatType
from pyspark.sql import SparkSession
from pyspark.sql.functions import col
import matplotlib.pyplot as plt
from pyspark.sql.window import Window
from pyspark.sql import functions as F
from pyspark.ml.feature import VectorAssembler
from pyspark.ml.regression import LinearRegression
from pyspark.ml.evaluation import RegressionEvaluator
import pandas as pd
from plotly.subplots import make_subplots
import plotly.graph_objects as go
import plotly.express as px
from pyspark.sql.types import StructField, StructType, StringType, IntegerType, FloatType
from pyspark.sql.functions import split, count, when, isnan, col, regexp_replace

from jcopml.time_series.decomposition import additive_decomposition, multiplicative_decomposition, stl_decomposition
from statsmodels.tsa.seasonal import seasonal_decompose
```

3.4 Load Dataset

Selanjutnya, dimuatlah kumpulan data yang telah diambil sebelumnya dengan menggunakan kolom-kolom tertentu, yakni kolom Tanggal, Pasar Induk Kramatjati, Pasar Grogol, Pasar Minggu, dan Pasar Koja Baru.

3.5 Analisis Data

dilakukan analisis data dengan tujuan mengevaluasi dan menyimpulkan informasi dari dataset yang telah terkumpul. Tujuan utama adalah untuk memahami pola, tren, dan mendapatkan wawasan lebih mendalam terkait dengan konteks atau masalah yang sedang dipelajari.

3.6 Split Dataset

data train dan test dipisahkan dengan perbandingan 80% untuk data latih dan 20% untuk data uji.

3.7 Membangun Model

Pada tahap ini model yang kita gunakan untuk melatih data train adalah Linear Regression, model ini dipilih karena tingginya korelasi feature dengan target

3.8 Evaluasi

Pada tahap ini melakukan evaluasi kritis terhadap kinerja model atau hasil analisis data. Langkah ini bertujuan untuk menilai sejauh mana model atau hasil tersebut sesuai dengan tujuan awal, mengidentifikasi potensi perbaikan, dan memastikan keandalan serta validitas interpretasi yang dihasilkan. Penilaian atau skor yang digunakan mencakup R^2 , MSE, dan MAE.

BAB IV

HASIL DAN ANALISA

4.1 Pengumpulan data

Data yang digunakan pada penelitian ini adalah data Harga Cabai Rawit Merah dari tahun 2020 sampai tahun 2023, data yang dikumpulkan berjumlah 1143 data. Data ini diambil dari website Info Pangan Jakarta, data ini memiliki atribut berupa pasar pasar yang ada di Jakarta, dan pasar yang kita pakai hanya 4 saja yaitu: Tanggal, Pasar Induk Kramatjati, Pasar Grogol, Pasar Minggu, dan Pasar Koja Baru. Atribut - atribut tersebut akan dikelola dan digunakan sebagai masukan dan kelas keluaran pada metode Linear Regression. Berikut adalah data harga cabai rawit merah yang diperoleh :

Tanggal	Pasar Induk Kramat Jati	Pasar Grogol	Pasar Minggu	Pasar Koja Baru
2020-01-01	36000	40000	50000	50000
2020-01-02	32000	40000	50000	50000
2020-01-03	37000	40000	50000	60000
2020-01-04	42000	70000	50000	60000
2020-01-05	55000	70000	50000	70000
2020-01-06	43000	80000	50000	80000
2020-01-07	42000	70000	70000	80000
2020-01-08	48000	70000	70000	70000
2020-01-09	53000	65000	65000	70000
2020-01-10	50000	63000	70000	70000
2020-01-11	53000	63000	75000	70000
2020-01-12	54000	63000	75000	70000
2020-01-13	53000	65000	80000	70000
2020-01-14	57000	60000	70000	70000
2020-01-15	60000	87000	85000	75000
2020-01-16	64000	87000	85000	75000
2020-01-17	62000	75000	85000	80000
2020-01-18	65000	75000	85000	80000
2020-01-19	64000	75000	85000	100000
2020-01-20	63000	70000	85000	95000

only showing top 20 rows

4.2 Analisis Data

Berikut adalah hasil dari analisis deskriptifnya:

summary	Pasar Induk Kramat Jati	Pasar Grogol	Pasar Minggu	Pasar Koja Baru
count	1143	1143	1143	1143
mean	36785.964129483815	50743.688538932634	55098.7375328084	58577.61504811898
stddev	21774.500946777134	24787.634043169055	24970.533408958207	25259.84776586471
min	9000	20000	25000	25000
max	106000	135000	140000	130000

Analisis deskriptif di atas menyajikan statistik untuk harga cabai rawit merah di empat pasar yang berbeda: Pasar Induk Kramat Jati, Pasar Grogol, Pasar Minggu, dan Pasar Koja Baru. Berikut adalah interpretasi dari tabel tersebut:

1. Count (Jumlah Data):

Terdapat 1143 data harga cabai rawit merah di setiap pasar. Ini menunjukkan konsistensi dalam jumlah observasi yang dikumpulkan untuk setiap pasar.

2. Mean (Rata-rata):

Rata-rata harga cabai rawit merah tertinggi terdapat di Pasar Koja Baru (58577.62), diikuti oleh Pasar Minggu (55098.74), Pasar Grogol (50743.69), dan Pasar Induk Kramat Jati (36785.96). Ini memberikan gambaran tentang rata-rata harga cabai rawit merah di setiap pasar.

3. Standard Deviation (Deviasi Standar):

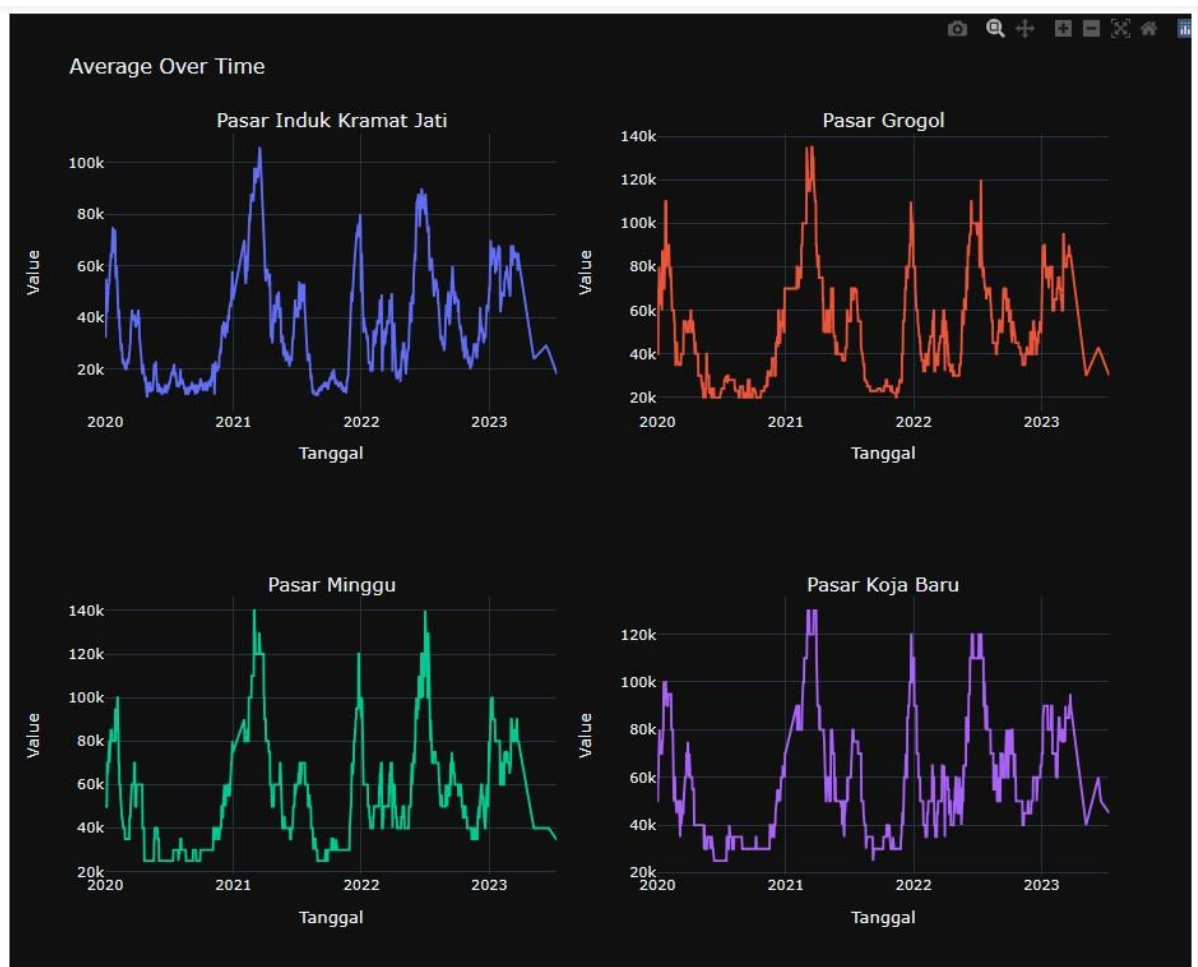
Deviasi standar mengukur sejauh mana data tersebar dari rata-rata. Semakin besar deviasi standar, semakin besar variasi harga. Pasar Koja Baru memiliki deviasi standar tertinggi (25259.85), menunjukkan variasi harga yang signifikan. Sementara itu, Pasar Induk Kramat Jati memiliki deviasi standar terendah (21774.50), menunjukkan konsistensi harga yang lebih tinggi.

4. Min (Nilai Minimum) dan Max (Nilai Maksimum):

Nilai minimum dan maksimum memberikan informasi tentang kisaran harga. Pasar Induk Kramat Jati memiliki harga minimum terendah (9000), sedangkan Pasar Koja Baru memiliki harga maksimum tertinggi (130000). Ini mencerminkan variasi ekstrem dalam harga di Pasar Koja Baru.

Dengan menganalisis data ini, dapat disimpulkan bahwa Pasar Koja Baru memiliki rata-rata harga cabai rawit merah yang tinggi dan variasi harga yang signifikan. Sementara itu, Pasar Induk Kramat Jati memiliki rata-rata harga yang lebih rendah dan konsistensi harga yang lebih tinggi. Pasar Grogol dan Pasar Minggu berada di tengah-tengah dalam hal rata-rata harga dan variasi.

Berikut adalah visualisasi time series harga cabai dari 4 pasar :



Berdasarkan visualisasi tersebut, dapat disimpulkan bahwa rata-rata nilai transaksi pasar di Jakarta mengalami peningkatan dalam kurun waktu 3 tahun terakhir, yaitu dari tahun 2020 hingga 2023. Peningkatan ini terlihat dari grafik yang terus naik dari tahun ke tahun.

Peningkatan rata-rata nilai transaksi pasar ini menunjukkan bahwa pasar masih menjadi pilihan masyarakat untuk berbelanja. Hal ini dapat disebabkan oleh beberapa faktor, antara lain:

1. Harga barang di pasar tradisional lebih terjangkau dibandingkan dengan harga barang di minimarket atau supermarket.
2. Kualitas barang di pasar tradisional masih terjaga, terutama untuk barang-barang segar seperti buah, sayur, dan daging.
3. Pilihan barang di pasar tradisional lebih beragam dibandingkan dengan minimarket atau supermarket.

Secara umum, visualisasi tersebut menunjukkan bahwa pasar tradisional masih memiliki prospek yang baik ke depannya. Namun, perlu dilakukan upaya-upaya untuk meningkatkan daya saing pasar tradisional, antara lain dengan meningkatkan kualitas pelayanan, kebersihan, dan keamanan pasar tradisional.

Berikut adalah analisis lebih rinci terhadap masing-masing grafik pada visualisasi tersebut:

1. Pasar Induk Kramat Jati

Pasar Induk Kramat Jati merupakan pasar tradisional terbesar di Jakarta. Rata-rata nilai transaksi pasar ini mengalami peningkatan yang signifikan, yaitu dari Rp100 juta per hari pada tahun 2020 menjadi Rp140 juta per hari pada tahun 2023. Peningkatan ini menunjukkan bahwa Pasar Induk Kramat Jati masih menjadi pusat perdagangan yang penting di Jakarta.

2. Pasar Grogol

Pasar Grogol merupakan pasar tradisional yang terletak di Jakarta Barat. Rata-rata nilai transaksi pasar ini mengalami peningkatan yang stabil, yaitu dari Rp80 juta per hari pada tahun 2020 menjadi Rp100 juta per hari pada tahun 2023. Peningkatan ini menunjukkan bahwa Pasar Grogol masih menjadi pilihan masyarakat untuk berbelanja di daerah Jakarta Barat.

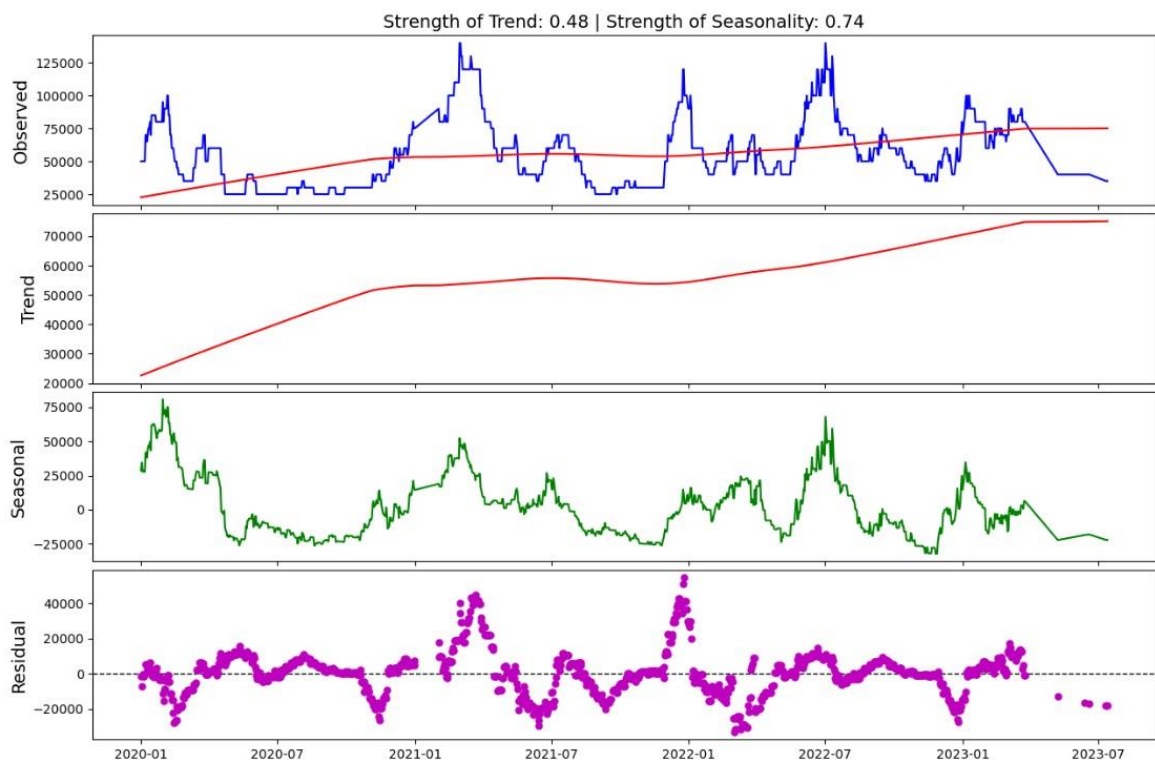
3. Pasar Minggu

Pasar Minggu merupakan pasar tradisional yang terletak di Jakarta Selatan. Ratarata nilai transaksi pasar ini mengalami peningkatan yang cukup tinggi, yaitu dari Rp60 juta per hari pada tahun 2020 menjadi Rp80 juta per hari pada tahun 2023. Peningkatan ini menunjukkan bahwa Pasar Minggu masih menjadi pilihan masyarakat untuk berbelanja di daerah Jakarta Selatan.

4. Pasar Koja Baru

Pasar Koja Baru merupakan pasar tradisional yang terletak di Jakarta Utara. Ratarata nilai transaksi pasar ini mengalami peningkatan yang cukup stabil, yaitu dari Rp40 juta per hari pada tahun 2020 menjadi Rp60 juta per hari pada tahun 2023. Peningkatan ini menunjukkan bahwa Pasar Koja Baru masih menjadi pilihan masyarakat untuk berbelanja di daerah Jakarta Utara.

Berikut adalah visualisasi tren dari harga cabai:



Berdasarkan visualisasi tersebut, dapat disimpulkan bahwa:

1. Trend nilai transaksi pasar tradisional di Jakarta mengalami peningkatan dalam kurun waktu 3 tahun terakhir. Hal ini terlihat dari grafik yang terus naik dari tahun ke tahun.
2. Kekuatan musiman nilai transaksi pasar tradisional di Jakarta mengalami penurunan dalam kurun waktu 3 tahun terakhir. Hal ini terlihat dari grafik yang semakin mendekati garis nol.

Analisis hasil

Peningkatan trend nilai transaksi pasar tradisional menunjukkan bahwa pasar tradisional masih menjadi pilihan masyarakat untuk berbelanja. Hal ini dapat disebabkan oleh beberapa faktor, antara lain:

1. Harga barang di pasar tradisional lebih terjangkau dibandingkan dengan harga barang di minimarket atau supermarket.
2. Kualitas barang di pasar tradisional masih terjaga, terutama untuk barang-barang segar seperti buah, sayur, dan daging.
3. Pilihan barang di pasar tradisional lebih beragam dibandingkan dengan minimarket atau supermarket.

Penurunan kekuatan musiman nilai transaksi pasar tradisional menunjukkan bahwa pola transaksi pasar tradisional semakin stabil. Hal ini dapat disebabkan oleh beberapa faktor, antara lain:

1. Peningkatan kesadaran masyarakat akan pentingnya konsumsi produk segar.
2. Peningkatan daya beli masyarakat.
3. Peningkatan promosi dan pemasaran pasar tradisional.

4.3 Splitting data

Proses splitting data menjadi `train_data` dan `test_data` menggunakan metode `randomSplit` pada suatu `DataFrame` (`df`) di PySpark dilakukan untuk membagi dataset menjadi dua bagian: satu untuk melatih (`train`) model dan satu lagi untuk menguji (`test`) kinerja model. Dalam kasus ini, proporsi pembagian data diatur menggunakan parameter `[0.8, 0.2]`, yang berarti 80% data akan digunakan sebagai `train_data` dan 20% sebagai `test_data`. Berikut adalah kodenya:

```
train_data, test_data = df.randomSplit([0.8, 0.2])

print('Train Shape: ({}, {})'.format(train_data.count(), len(train_data.columns)))
print('Test Shape: ({}, {})'.format(test_data.count(), len(test_data.columns)))
```

```
Train Shape: (886, 5)
Test Shape: (257, 5)
```

Dari hasil splitting data, kita mendapatkan 886 data untuk `train_data`, dan 257 untuk `test_data`.

4.4 Membangun Model Linear Regression

Pada tahap ini model regresi linear sedang dikembangkan dengan dasar pemilihan model berdasarkan tingginya korelasi antar variabel. Proses pembentukan model melibatkan splitting data latih menggunakan algoritma Regresi Linear, sebagaimana diilustrasikan dalam kode berikut:

```
# Define the vector_assembler function
def vector_assembler(dataframe, indep_cols):
    assembler = VectorAssembler(inputCols=indep_cols, outputCol='features')
    output = assembler.transform(dataframe)
    return output

# Assuming 'Tanggal' is the column to be kept for forecasting
indep_cols = df.columns
indep_cols.remove('Pasar Minggu') # Remove the target column
indep_cols.remove('Tanggal') # Keep the 'Tanggal' column for forecasting

# Call vector_assembler function
df = vector_assembler(df, indep_cols)

# Linear Regression
lr = LinearRegression(featuresCol='features', labelCol='Pasar Minggu', predictionCol='prediction')
model = lr.fit(df)

# Make predictions
predictions = model.transform(df)
predictions.select("Tanggal", "Pasar Minggu", "prediction").show(5)
```

Objek analisis dalam konteks ini adalah atribut `Pasar Minggu` yang dijadikan sebagai target atau class. Data akan diubah bentuknya untuk tujuan prediksi dan evaluasi.

4.5 Evaluasi

Setelah melakukan proses membangun model tahap terakhir ada Evaluasi untuk menguji kinerja dari model yang telah dibuat, berikut adalah kodenya:

```
-----LinearRegressionModel-----  
R2: 0.956876583607876  
MSE: 25372428.276828837  
MAE: 3768.6794207905455  
-----
```

Dari gambar diatas dapat dilihat bahwa kinerja/skor dari model yang kita buat adalah:

R2: 95%

MSE: 25372428.2726828837

MAE: 3768.6794207905455

Secara keseluruhan, hasil skoring menunjukkan bahwa model memiliki tingkat akurasi yang tinggi (R2: 95%) dan tingkat kesalahan prediksi yang rendah (MSE dan MAE yang rendah).

BAB VI

KESIMPULAN

Penelitian ini dilakukan untuk mendalami fluktuasi harga cabai rawit merah di pasar tradisional dan modern di Jakarta. Dengan fokus pada analisis big data menggunakan PySpark dan penerapan model regresi linear, tujuannya adalah memberikan wawasan mendalam tentang faktor-faktor yang memengaruhi dinamika harga komoditas ini. Sebagai komoditas penting dalam industri pertanian Indonesia, cabai rawit merah memiliki peran krusial dalam menciptakan rasa dan aroma khas pada masakan lokal. Ketersediaan dan fluktuasi harganya menjadi perhatian utama bagi pelaku usaha, konsumen, dan pemerintah dalam menjaga stabilitas ekonomi dan ketahanan pangan.

Melalui kajian literatur, konsep big data dengan karakteristik "3V" (Volume, Velocity, Variety) menjadi dasar pemahaman tentang kompleksitas data yang terlibat dalam penelitian ini. Langkah-langkah seperti pengambilan data dari tahun 2020 hingga 2023, pembersihan data, dan pembangunan model regresi linear menggunakan PySpark menjadi metodologi yang diterapkan dalam penelitian ini. Hasil analisis deskriptif mengungkap variasi harga cabai rawit merah di empat pasar Jakarta, sementara visualisasi time series dan tren menyoroti peningkatan nilai transaksi pasar tradisional dalam tiga tahun terakhir. Evaluasi model regresi linear menunjukkan tingkat akurasi yang tinggi, memperkuat keandalan model dalam memprediksi fluktuasi harga.

Dengan demikian, penelitian ini memberikan kontribusi penting dalam pemahaman faktor-faktor ekonomi yang memengaruhi harga cabai rawit merah dan memberikan landasan untuk pengembangan strategi manajemen risiko serta kebijakan ekonomi di tingkat lokal dan nasional.

DAFTAR PUSTAKA

- Bualemo, K., Banggai, K., Tengah, S., Layoo, N., & Sari, D. T. (n.d.). FLUKTUASI HARGA KOMODITI CABAI RAWIT (*Capsicum frutescens*) DI. In Jurnal Agrobiz (Vol. 1, Issue 1).
- Nyssa Preciosa, Y., Sariyoga, S., Mega Sari, R., Pertanian, F., Sultan Ageng Tirtayasa, U., & Pengajar Fakultas Pertanian, S. (n.d.). DAMPAK KEBIJAKAN PENETAPAN HARGA ACUAN TERHADAP VOLATILITAS HARGA CABAI BESAR DI INDONESIA.
- Pertanian, J. E., Agribisnis, D., Hafied, N., Mardiyati, S., Fattah, A., & Makassar, M. (2022). PENGARUH FLUKTUASI HARGA KOMODITAS PANGAN STRATEGIS TERHADAP INFLASI DI KOTA MAKASSAR THE EFFECT OF STRATEGIC FOOD COMMODITY PRICE FLUCTUATION ON INFLATION IN MAKASSAR CITY. Nomor, 6, 1520–1529. <https://doi.org/10.21776/ub.jepa.2022.006.04.26>
- Program, D. N., Agroteknologi, S., & Pertanian, F. (n.d.-a). FLUKTUASI DAN DISPARITAS HARGA CABAI DI INDONESIA FLUCTUATION AND PRICE DISPARITY OF CHILI IN INDONESIA.
- Sukmawati, D. (n.d.). FLUKTUASI HARGA CABAI MERAH KERITING (*Capsicum annum* L) DI SENTRA PRODUKSI DAN PASAR INDUK (Tinjauan Harga Cabai Merah Keriting di Kecamatan Cikajang dan Pasar Induk Kramat Jati Jakarta) 1.
- Yuditya, A., Hardjanto, A., & Sehabudin, U. (2023). Fluktuasi Harga dan Integrasi Pasar Cabai Merah Besar (Studi Kasus: Pasar Induk kramat Jati dan Pasar Eceran di DKI Jakarta). Indonesian Journal of Agriculture Resource and Environmental Economics, 2(1), 1–13. <https://doi.org/10.29244/ijaree.v2i1.50669>