

TEKNOLOGI BIG DATA
LAPORAN CRIME DATA ANALYSIS



Disusun oleh :

Kelompok 5

Daffa Anis Fahrizi	(2106731365)
Miranti Anggunsari	(2106731472)
Lavly Rantissa Zunnuraina R.	(2206830624)
Safia Amita Khoirunnisa	(2206059420)
Annisa Ardelia Setiawan	(2206059471)

Fakultas Teknik, Departemen Teknik Elektro

Program Studi Teknik Komputer

Universitas Indonesia

Tahun Ajaran 2024/2025

Daftar Isi

Daftar Isi.....	1
Bab I Pendahuluan.....	2
1.1. Latar Belakang.....	2
1.2. Rumusan Masalah.....	3
1.3. Tujuan.....	3
Bab II Dasar Teori.....	4
2.1. Hadoop.....	4
2.2. Apache Spark.....	5
2.3. Matplotlib, Pandas, dan Seaborn.....	6
Bab III Implementasi.....	7
3.1. Data Cleaning.....	7
3.2. Data Processing.....	10
3.3. Data Visualization.....	14
Bab IV Pembahasan.....	18
4.1. Tren Kejahatan Berdasarkan Tahun.....	18
4.2. Jenis Kejahatan yang Paling Sering Terjadi.....	20
4.3. Distribusi Kejahatan Berdasarkan Wilayah.....	21
4.3.1. Wilayah dengan Potensi Kejahatan Tinggi.....	21
4.3.2. Wilayah dengan Potensi Kejahatan Rendah.....	22
4.3.3. Pola Unik Berdasarkan Wilayah.....	22
4.4. Analisis Kombinasi.....	23
4.4.1. Pencurian di Cambridgeport dan East Cambridge.....	23
4.4.2. Vandalisme di Riverside dan Mid-Cambridge.....	23
4.4.3. Kejahatan Berat yang Tersebar Secara Merata.....	24
4.5. Tren Sosial dan Kriminalitas.....	25
4.6. Solusi dan Rekomendasi.....	26
4.7. Kemungkinan Analisis Lanjutan.....	29
4.7.1. Clustering Wilayah Berdasarkan Jenis Kejahatan.....	29
4.7.2. Prediksi Tren dengan Time Series Analysis.....	29
4.7.3. Analisis Korelasi Sosio-Ekonomi.....	30
Bab V Kesimpulan.....	31
Referensi.....	32

Bab I

Pendahuluan

1.1. Latar Belakang

Di tengah pesatnya pertumbuhan ekonomi dan teknologi, tentunya kejahatan juga ikut berkembang dengan berbagai bentuk dan modus operandi yang semakin kompleks. Dalam beberapa dekade terakhir, urbanisasi yang cepat dan ketimpangan sosial telah menjadi pemicu utama peningkatan tingkat kriminalitas. Lingkungan perkotaan yang padat sering kali menjadi pusat berbagai jenis kejahatan, mulai dari pencurian, perampokan, hingga tindak kekerasan.

Menghadapi situasi ini, pendekatan tradisional dalam penanganan kejahatan sering kali kurang efektif tanpa adanya dukungan data yang akurat dan komprehensif. Oleh karena itu, analisis data kriminalitas menjadi sangat penting untuk memahami dinamika kejahatan di suatu wilayah.

Analisis ini kami lakukan dengan tujuan untuk mengeksplorasi pola kejahatan berdasarkan dataset kejahatan di Cambridge, Massachusetts. Dengan memanfaatkan data historis mengenai laporan kejahatan, analisis ini diharapkan dapat memberikan wawasan mengenai tren kriminalitas, lokasi yang rawan kejahatan, dan jenis kejahatan yang dominan. Hasil dari analisis ini dapat digunakan sebagai dasar untuk meningkatkan strategi keamanan dan menciptakan lingkungan yang lebih aman bagi masyarakat.

Untuk menangani dataset berukuran besar ini, kami memilih menggunakan dua tools utama, yaitu Hadoop dan Apache Spark. Hadoop menyediakan fitur-fitur yang memungkinkan pengolahan dan penyimpanan data dalam skala besar secara efisien. Sementara itu, Apache Spark, sebagai mesin pemrosesan data open-source, menawarkan kecepatan komputasi yang tinggi serta kemampuan skalabilitas untuk memenuhi tuntutan pengolahan data besar. Dengan mengkombinasikan kedua tools ini, kami berharap dapat menjalankan analisis data dengan lebih efisien dan akurat.

Selain Hadoop dan Apache Spark, kami juga memanfaatkan Apache Zeppelin, sebuah notebook berbasis aplikasi web, untuk meningkatkan interaktivitas dalam analisis data. Alat ini memungkinkan kami melakukan kueri

data tanpa perlu membuat skema secara manual, sehingga mempermudah proses visualisasi data. Dengan fitur-fitur yang ditawarkan, Apache Zeppelin mendukung pembuatan visualisasi yang lebih efektif dan interaktif, mempercepat pemahaman terhadap data yang dianalisis.

1.2. Rumusan Masalah

Berikut ini rumusan masalah yang mendasari laporan analisis ini:

1. Bagaimana distribusi geografis dan temporal kejahatan di berbagai lingkungan perkotaan di Cambridge?
2. Jenis kejahatan apa yang paling dominan terjadi, dan di wilayah mana kejahatan tersebut paling sering terjadi?
3. Berdasarkan data historis yang ada, dalam satu tahun pada bulan apa saja tren kejahatan mengalami peningkatan?

1.3. Tujuan

Berikut ini tujuan yang mendasari pembentukan laporan analisis ini:

1. Mengidentifikasi pola dan tren kejahatan di Cambridge untuk memahami dinamika kriminalitas di wilayah tersebut.
2. Menganalisis distribusi spasial dan temporal kejahatan guna mengidentifikasi area-area rawan serta waktu dengan tingkat kejadian tinggi.
3. Mengklasifikasikan jenis kejahatan yang dominan dan memahami karakteristik masing-masing jenis kejahatan berdasarkan data historis.

Bab II

Dasar Teori

2.1. Hadoop



Hadoop adalah framework open-source yang dikembangkan oleh Apache Software Foundation untuk penyimpanan dan pemrosesan data dalam skala besar secara terdistribusi. Hadoop dirancang untuk menangani data dalam jumlah besar (big data) yang tidak dapat diproses secara efisien menggunakan metode tradisional. Hadoop menggunakan konsep penyimpanan terdistribusi dan pemrosesan paralel, sehingga memungkinkan pengelolaan data yang efisien dan cepat.

Komponen utama Hadoop:

- **Hadoop Distributed File System (HDFS):** HDFS adalah sistem penyimpanan terdistribusi yang memungkinkan file besar dipecah menjadi blok-blok kecil dan disimpan di berbagai node dalam kluster. Hal ini memberikan keandalan tinggi karena data tetap dapat diakses meskipun salah satu node gagal.
- **MapReduce:** Model pemrograman yang mendukung pemrosesan paralel dengan membagi tugas menjadi dua tahap utama, yaitu Map (proses pemetaan data mentah ke format yang dapat digunakan) dan Reduce (proses agregasi hasil map untuk menghasilkan output akhir).
- **YARN (Yet Another Resource Negotiator):** Framework untuk manajemen sumber daya dalam *cluster* Hadoop yang mengalokasikan CPU dan memori untuk aplikasi yang berjalan.

Dalam proyek ini, Hadoop digunakan untuk menyimpan dataset besar secara terdistribusi. Dataset kejahatan memiliki ukuran besar dan membutuhkan sistem penyimpanan yang andal dan efisien. Dengan menggunakan HDFS, dataset dapat diakses secara cepat dan aman oleh berbagai proses pemrosesan data seperti cleaning dan analisis yang dilakukan menggunakan Apache Spark.

2.2. Apache Spark



Apache Spark adalah framework open-source untuk komputasi kluster yang dirancang untuk kecepatan, kemudahan penggunaan, dan analisis data real-time. Spark merupakan salah satu framework big data yang paling populer karena mendukung pemrosesan

data besar dengan kecepatan tinggi.

Fitur utama Apache Spark:

- **In-Memory Computing:** Spark memproses data dengan menyimpannya di memori, sehingga menghindari biaya I/O disk yang tinggi seperti yang terjadi pada MapReduce.
- **Library Terintegrasi:** Spark menyediakan berbagai library bawaan, seperti:
 - **Spark SQL:** Untuk pemrosesan data terstruktur menggunakan bahasa SQL.
 - **MLlib:** Untuk analisis dan penerapan machine learning.
 - **GraphX:** Untuk analisis data berbasis graf.
- **Dukungan Multi-Language:** Mendukung berbagai bahasa pemrograman seperti Python, Scala, R, dan Java, sehingga fleksibel untuk berbagai kebutuhan pengembang.
- **Kompatibilitas dengan Hadoop:** Spark dapat dijalankan di atas Hadoop dan memanfaatkan HDFS untuk penyimpanan data.

Secara khusus, kami memilih modul SparkSQL untuk melakukan pengolahan dataset. SparkSQL memberikan pendekatan yang lebih terfokus pada pengolahan data terstruktur menggunakan SQL di dalam ekosistem Apache Spark. SparkSQL sendiri merupakan modul dalam Apache Spark yang memungkinkan pemrosesan data terstruktur dan semi-terstruktur menggunakan bahasa kueri SQL. Hal ini memungkinkan analisis yang lebih intuitif dan efisien ketika data yang digunakan memiliki struktur tabel, seperti data kriminal yang biasanya tersimpan dalam format CSV atau database SQL. Hasil dari kueri SparkSQL ini dapat langsung dianalisis lebih lanjut menggunakan visualisasi dengan Matplotlib.

2.3. Matplotlib, Pandas, dan Seaborn

Matplotlib, Pandas, dan Seaborn adalah library visualisasi data yang sangat populer dan fleksibel dalam ekosistem Python. Library ini memungkinkan kami untuk membuat berbagai jenis grafik dan plot secara kustom, yang sangat berguna untuk analisis data. Dengan Matplotlib, kami dapat mempresentasikan data dalam bentuk visual yang informatif, memperjelas pola dan tren yang mungkin tidak terlihat dalam bentuk data mentah.



Fitur utama Matplotlib:

- **Multi-plot:** Mendukung pembuatan berbagai jenis plot, termasuk grafik garis (line plot), batang (bar plot), sebaran (scatter plot), histogram, dan lainnya. Fleksibilitas ini memungkinkan pengguna untuk menyesuaikan visualisasi sesuai dengan kebutuhan analisis data tertentu.
- **Kustomisasi Lengkap:** Memiliki kemampuan untuk mengontrol setiap aspek visualisasi, mulai dari warna, label sumbu, judul, legenda, hingga anotasi data. Fitur ini memberikan kebebasan bagi analis data untuk menciptakan visualisasi yang sesuai dengan kebutuhan spesifik.
- **Integrasi dan Metode Ekspor yang Lengkap:** Matplotlib mendukung penyimpanan grafik dalam berbagai format seperti PNG, PDF, SVG, dan lainnya. Grafik yang dibuat dapat dengan mudah diintegrasikan ke dalam laporan, presentasi, atau aplikasi web. Selain itu, Matplotlib bekerja dengan baik bersama library Python lainnya seperti NumPy dan Pandas, memungkinkan integrasi yang mulus dalam alur kerja analisis data.

Dalam proyek ini, Matplotlib digunakan untuk membuat visualisasi distribusi kejahatan berdasarkan tahun, jenis kejahatan, dan wilayah (wilayah). Dengan Matplotlib, data dari Hadoop dan Spark yang telah diproses dapat divisualisasikan menjadi grafik garis dan batang yang memberikan wawasan lebih dalam. Fitur anotasi dan kustomisasi Matplotlib membantu menyoroti poin-poin penting, seperti jumlah kejahatan di setiap tahun atau perbedaan mencolok antara jenis kejahatan. Dengan menyimpan visualisasi dalam format gambar, hasil analisis dapat dengan mudah didokumentasikan dan dibagikan.

Bab III

Implementasi



CrimeAnalysis_FeatureEngineering

Updated 5mo ago

[6 comments](#) · Cambridge Crime Data

▲ 34

🏅 Bronze ...

Dalam proyek ini, kami menggunakan dataset yang sudah didapatkan dari *kaggle*. Implementasi proyek ini dibagi menjadi tiga tahap utama, yaitu *data cleaning*, *data processing*, dan *data visualization*. Berikut adalah penjelasan detail untuk setiap tahap implementasi.

3.1. Data Cleaning

Dalam kasus ini, kami menggunakan Apache Spark untuk membersihkan dataset laporan kejahatan, dengan integrasi ke Hadoop untuk penyimpanan data besar. Proses data cleaning ini memastikan bahwa dataset kejahatan bersih, terorganisir, dan siap untuk analisis mendalam. Dengan menghilangkan kesalahan data dan menyusun informasi penting, seperti tahun kejadian, analisis tren kriminalitas dapat dilakukan dengan lebih akurat dan dapat diandalkan. Sebelumnya, dataset yang masih kotor akan disimpan ke dalam HDFS.

Browse Directory

/data

Go!

Search:

Show

25

entries

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 20:58	0	0 B	cleaned_crime_reports1.csv	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 21:04	0	0 B	cleaned_crime_reports2.csv	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 17:54	0	0 B	crime_reports	<input type="checkbox"/>

Showing 1 to 3 of 3 entries

Previous

1

Next

Hadoop, 2024.

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import regexp_extract, col

# 1. Inisialisasi Spark Session
spark = SparkSession.builder \
    .appName("Crime Data Cleaning") \
    .config("spark.executor.memory", "2g") \
```



```

.config("spark.driver.memory", "2g") \
    .getOrCreate()

# 2. Memuat Dataset dari HDFS
hdfs_path = "hdfs:///data/crime_reports/crime_reports.csv" #
# Path dataset asli
df = spark.read.csv(hdfs_path, header=True, inferSchema=True)

# Data sebelum pembersihan
print(f"Jumlah baris sebelum pembersihan: {df.count()}")

# 3. Menghapus baris dengan nilai "Admin Error"
df_clean = df.filter(~(
    (df['Crime'] == 'Admin Error') |
    (df['Wilayah'] == 'Admin Error') |
    (df['Location'] == 'Admin Error')
))

# 4. Mengganti nilai kosong (None) dengan "NA"
df_clean = df_clean.fillna("NA")

# 5. Ekstrak tahun dari kolom 'Crime Date Time'
df_clean = df_clean.withColumn(
    "Crime Year",
    regexp_extract(col("Crime Date Time"), r"(\d{4})", 1) #
    Ambil tahun pertama yang ditemukan
)

# Verifikasi data setelah pembersihan
print(f"Jumlah baris setelah pembersihan: {df_clean.count()}")
df_clean.show(5)

# 6. Simpan dataset bersih ke HDFS dalam satu file
output_clean_path = "hdfs:///data/cleaned_crime_reports2.csv"
# Path dataset bersih
df_clean.coalesce(1).write.csv(output_clean_path, header=True)

# 7. Menutup Spark session
spark.stop()

```

Kode di atas bertujuan untuk membersihkan dataset mentah dari HDFS dan mempersiapkannya untuk analisis lebih lanjut. Proses dimulai dengan

inisialisasi `SparkSession` sebagai entry point utama untuk menggunakan Spark SQL, di mana sesi ini diberikan nama "Crime Data Cleaning" dan diatur dengan konfigurasi seperti alokasi memori sebesar 2 GB untuk driver dan eksekutor. Dataset kemudian dimuat dari lokasi file yang ditentukan di HDFS menggunakan perintah `spark.read.csv()` dengan opsi `header=True` untuk mengenali header pada file dan `inferSchema=True` agar tipe data kolom diidentifikasi secara otomatis. Dataset yang dimuat akan disimpan dalam sebuah `DataFrame Spark`.

Sebelum proses pembersihan, kode menghitung jumlah total baris menggunakan `df.count()` untuk mengetahui jumlah data awal. Data kemudian difilter untuk menghapus baris yang memiliki nilai "Admin Error" pada kolom Crime, Neighborhood, atau Location dengan menggunakan `filter()` dan operator logika seperti `~` (NOT) dan `|` (OR). Nilai kosong pada dataset akan diganti dengan "NA" menggunakan `fillna()` untuk memastikan data lebih konsisten. Selanjutnya, kolom baru bernama Crime Year akan ditambahkan dengan mengekstraksi informasi tahun dari kolom Crime Date Time menggunakan ekspresi reguler dalam `regexp_extract()`.

Hasil pembersihan ini diperiksa kembali dengan menghitung jumlah baris setelah pembersihan menggunakan `df_clean.count()` dan menampilkan lima baris pertama dataset yang telah bersih menggunakan `df_clean.show(5)`. Dataset bersih ini kemudian disimpan kembali ke HDFS dalam format CSV menggunakan `write.csv()` dengan opsi `header=True`, di mana partisi data akan digabungkan menjadi satu file menggunakan `coalesce(1)`. Terakhir, sesi Spark ditutup dengan perintah `spark.stop()` untuk membebaskan sumber daya. Berikut ini output dari proses *data cleaning* yang telah dilakukan:

```
daffafahriz1@daffafahriz12:~/Downloads/proyek akhir big data$ spark-submit cleaningData.py
24/12/04 14:03:56 WARN Utils: Your hostname, daffafahriz12 resolves to a loopback address: 127.0.1.1; using 10.0.2.15 instead (on interface enp0s3)
24/12/04 14:03:56 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
24/12/04 14:03:57 INFO SparkContext: Running Spark version 3.5.3
24/12/04 14:03:57 INFO SparkContext: OS info Linux, 6.8.0-49-generic, amd64
24/12/04 14:03:57 INFO SparkContext: Java version 1.8.0_432
24/12/04 14:03:58 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

File Number	Date of Report	Crime Date Time	Crime	Reporting Area	Neighborhood	Location	Crime Year
2009-01323	02/21/2009 09:53:...	02/21/2009 09:20 ...	Threats	105	East Cambridge	100 OTIS ST, Camb...	2009
2009-01324	02/21/2009 09:59:...	02/20/2009 22:30 ...	Auto Theft	1109	North Cambridge	400 RINDGE AVE, C...	2009
2009-01327	02/21/2009 12:32:...	02/19/2009 21:00 ...	Hit and Run	1109	North Cambridge	400 RINDGE AVE, C...	2009
2009-01331	02/21/2009 03:05:...	02/21/2009 15:00 ...	Larceny (Misc)	1303	Strawberry Hill	0 NORUMBEGA ST, C...	2009
2009-01346	02/22/2009 05:02:...	02/22/2009 05:02	OUI	105	East Cambridge	FIFTH ST & GORE S...	2009

only showing top 5 rows

Setelah dibersihkan, data-data ini akan disimpan kembali ke dalam HDFS dalam satu folder baru yang dikhususkan untuk data yang bersih. Dataset awal atau kotor awalnya berjumlah 95.924 dan setelah dilakukan *data cleaning*, data bersih yang diperoleh berkurang menjadi sebanyak 92.755.

Browse Directory

/data

Go!

Show

25

entries

Search:

<input type="checkbox"/>	<div>Permission</div>	<div>Owner</div>	<div>Group</div>	<div>Size</div>	<div>Last Modified</div>	<div>Replication</div>	<div>Block Size</div>	<div>Name</div>	<div></div>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 20:58	0	0 B	cleaned_crime_reports1.csv	<div></div>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 21:04	0	0 B	cleaned_crime_reports2.csv	<div></div>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 17:54	0	0 B	crime_reports	<div></div>

Showing 1 to 3 of 3 entries

Previous

1

Next

Hadoop, 2024.

3.2. Data Processing

Setelah data dibersihkan, proses analisis data dilakukan untuk mengidentifikasi pola dan tren yang relevan dengan kasus kejahatan. Pemrosesan data dilakukan menggunakan Apache Spark, yang memungkinkan pengolahan data besar dengan efisiensi tinggi. Tiga jenis agregasi dilakukan, yaitu jumlah kejahatan per tahun untuk analisis temporal, jumlah kejahatan per jenis untuk memahami kejahatan dominan, dan jumlah kejahatan berdasarkan wilayah (wilayah) untuk mengidentifikasi area rawan. Data hasil agregasi kemudian disimpan di HDFS sebagai file terpisah, yang selanjutnya digunakan untuk visualisasi.

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import col

# 1. Inisialisasi Spark Session
spark = SparkSession.builder \
    .appName("Analyze Crime Data by Year") \
    .config("spark.executor.memory", "2g") \
    .config("spark.driver.memory", "2g") \
    .getOrCreate()
```

```

# 2. Memuat Dataset yang Sudah Dibersihkan
cleaned_path = "hdfs:///data/cleaned_crime_reports2.csv"
df_cleaned = spark.read.csv(cleaned_path, header=True,
inferSchema=True)

# 3. Hitung jumlah kejahatan per tahun
annual_crimes = df_cleaned.groupBy("Crime
Year").count().orderBy("Crime Year")

# 4. Hitung jumlah kejahatan per Wilayah
wilayah_crimes =
df_cleaned.groupBy("Wilayah").count().orderBy("count",
ascending=False)

# 5. Hitung jumlah kejahatan per Crime Type
crime_type_crimes =
df_cleaned.groupBy("Crime").count().orderBy("count",
ascending=False)

# 6. Tampilkan hasil
print("Jumlah Kejahatan per Tahun:")
annual_crimes.show()

print("Jumlah Kejahatan per Wilayah:")
wilayah_crimes.show()

print("Jumlah Kejahatan per Crime Type:")
crime_type_crimes.show()

# 7. Simpan hasil pengolahan ke HDFS
output_base_path = "hdfs:///output/Crime11"
annual_crimes.write.csv(f"{output_base_path}/annual_crimes.csv
", header=True)
wilayah_crimes.write.csv(f"{output_base_path}/wilayah_crimes.c
sv", header=True)
crime_type_crimes.write.csv(f"{output_base_path}/crime_type_cr
imes.csv", header=True)

# 8. Menutup Spark session
spark.stop()

```

Kode di atas melakukan pemrosesan dataset yang telah dibersihkan untuk menghasilkan laporan analitik berdasarkan tahun, lokasi, dan jenis kejahatan. Proses ini dilakukan dengan menggunakan Spark untuk menangani data berskala besar.

```
daffafahrizi@daffafahrizi2:~/Downloads/proyek_akhir_big_data$ spark-submit processingData.py
24/12/04 15:26:45 WARN Utils: Your hostname, daffafahrizi2 resolves to a loopback address: 127.0.1.1; using 10.0.2.15 instead (on interface enp0s3)
24/12/04 15:26:45 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
24/12/04 15:26:46 INFO SparkContext: Running Spark version 3.5.3
24/12/04 15:26:46 INFO SparkContext: OS info Linux, 6.8.0-49-generic, amd64
24/12/04 15:26:46 INFO SparkContext: Java version 1.8.0_432
24/12/04 15:26:46 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
24/12/04 15:26:46 INFO ResourceUtils: =====
24/12/04 15:26:46 INFO ResourceUtils: No custom resources configured for spark.driver.
24/12/04 15:26:46 INFO ResourceUtils: =====
```

Pertama-tama, `SparkSession` akan diinisialisasi dan diberikan nama untuk memastikan bahwa sesi Spark telah tersedia dan siap digunakan. Kemudian, dengan perintah `cleaned_path`, program akan diarahkan ke lokasi dataset bersih di HDFS. Dan program akan membaca dataset CSV yang kemudian akan dimuat untuk proses analisis.

Setelah itu, data akan dikelompokkan menjadi tiga tahapan, yaitu pengelompokkan data berdasarkan kolom *Crime Year*, kolom *Neighborhood*, serta kolom *Crime*. Dari masing-masing tahapan ini nantinya akan menghasilkan hasil analisis dalam format tabel yang menampilkan jumlah kejahatan per tahun, per wilayah, dan berdasarkan jenis kejahatannya.

Lalu, program akan melanjutkan proses ke penyimpanan hasil analisis ke dalam lokasi folder output dasar di HDFS dengan menggunakan perintah `output_base_path`. Kemudian, dengan perintah `write.csv()`, data tersebut akan disimpan dalam format CSV, di mana masing-masing file akan disimpan dalam folder yang berbeda berdasarkan kategorinya. Setelah data berhasil tersimpan, program akan menutup sesi Spark untuk membebaskan sumber daya dengan menggunakan perintah `spark.stop()`.

Browse Directory

/

Go!

Show

25

entries

Search:

<input type="checkbox"/>		Permission		Owner		Group		Size		Last Modified		Replication		Block Size		Name	
<input type="checkbox"/>		drwxr-xr-x		daffafahrizi		supergroup		0 B		Dec 04 21:04		0		0 B		data	
<input type="checkbox"/>		drwxr-xr-x		daffafahrizi		supergroup		0 B		Dec 04 22:27		0		0 B		output	

Showing 1 to 2 of 2 entries

Previous

1

Next

Hadron, 2024

Hadoop, 2024.

Browse Directory

/output

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 18:31	0	0 B	Crim9	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 18:41	0	0 B	Crime10	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 22:27	0	0 B	Crime11	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 15:34	0	0 B	Crime3	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 15:37	0	0 B	Crime4	<input type="checkbox"/>

Browse Directory

/output/Crime11

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 22:27	0	0 B	annual_crimes.csv	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 22:27	0	0 B	crime_type_crimes.csv	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	daffafahrizi	supergroup	0 B	Dec 04 22:27	0	0 B	neighborhood_crimes.csv	<input type="checkbox"/>

Showing 1 to 3 of 3 entries

Previous 1 Next

Hadoop, 2024.

Berikut ini output dari *data processing* yang sudah dilakukan dalam format tabel:

Jumlah Kejahatan per Tahun

```
24/12/04 15:26:58 INFO CodeGenerator: Code generated in 22.127547 ms
+-----+
|Crime Year|count|
+-----+
|      1980|    2|
|      1990|    1|
|      1993|    2|
|      1995|    1|
|      1999|    1|
|      2000|    5|
|      2001|   14|
|      2002|    4|
|      2003|    2|
|      2004|   16|
|      2005|    8|
|      2006|   17|
|      2007|   25|
|      2008|  121|
|      2009| 6515|
|      2010| 6474|
|      2011| 6433|
|      2012| 6144|
|      2013| 6285|
|      2014| 6179|
+-----+
only showing top 20 rows
```

Jumlah Kejahatan per Wilayah	Jumlah Kejahatan per Jenis
<pre> +-----+-----+ Neighborhood count +-----+-----+ Cambridgeport 14080 East Cambridge 11898 North Cambridge 9733 Mid-Cambridge 9549 Area 4 9236 Riverside 9211 West Cambridge 9028 Peabody 5824 Inman/Harrington 5617 Highlands 2571 Agassiz 2533 MIT 1939 Strawberry Hill 1536 +-----+-----+ </pre>	<pre> +-----+-----+ Crime count +-----+-----+ Hit and Run 9121 Larceny from MV 7369 Larceny of Bicycle 6292 Mal. Dest. Property 6130 Forgery 6109 Shoplifting 5610 Larceny from Bul... 4464 Warrant Arrest 4405 Simple Assault 4201 Larceny from Resi... 3953 Housebreak 3804 Larceny from Person 3228 Accident 2871 Threats 2766 Flim Flam 2583 Aggravated Assault 2399 Missing Person 1953 Auto Theft 1941 Harassment 1521 Street Robbery 1223 +-----+-----+ only showing top 20 rows </pre>

3.3. Data Visualization

Data visualization adalah tahap akhir untuk mempresentasikan hasil analisis data dalam bentuk yang lebih mudah dipahami. Dalam proyek ini, Matplotlib dan Seaborn digunakan untuk menghasilkan grafik seperti line chart untuk menunjukkan *annual crimes trend*, serta bar chart untuk distribusi jenis kejahatan dan wilayah rawan. Visualisasi membantu mengkomunikasikan hasil analisis kepada pembaca atau pengambil keputusan dengan lebih efektif. Misalnya, grafik tren tahunan menunjukkan lonjakan kejahatan pada 2009 dan penurunan drastis pada 2024, sementara grafik distribusi wilayah mengidentifikasi Cambridge Port sebagai area dengan tingkat kejahatan tertinggi.

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from pyspark.sql import SparkSession

# Inisialisasi Spark session
spark = SparkSession.builder.appName('CrimeDataAnalysis').getOrCreate()

```

```

# Membaca data dari Hadoop/HDFS
annual_crimes =
spark.read.csv('hdfs:///output/Crime11/annual_crimes.csv/part-
00000.csv', header=True, inferSchema=True)
crime_type_crimes =
spark.read.csv('hdfs:///output/Crime11/crime_type_crimes.csv/p
art-00000.csv', header=True, inferSchema=True)
wilayah_crimes =
spark.read.csv('hdfs:///output/Crime11/wilayah_crimes.csv/part
-00000.csv', header=True, inferSchema=True)

# Convert Spark DataFrame to pandas DataFrame untuk
visualisasi
annual_crimes_df = annual_crimes.toPandas()
crime_type_crimes_df = crime_type_crimes.toPandas()
wilayah_crimes_df = wilayah_crimes.toPandas()

# 1. Visualisasi Data Kejahatan Tahunan (Annual Crimes)
plt.figure(figsize=(10, 6))
sns.lineplot(x='Crime Year', y='count', data=annual_crimes_df,
marker='o', color='b')
plt.title('Tren Kejahatan Tahunan')
plt.xlabel('Tahun')
plt.ylabel('Jumlah Kejahatan')
plt.grid(True)
plt.xticks(rotation=45)
for i in range(len(annual_crimes_df)):
    plt.text(annual_crimes_df['Crime Year'][i],
annual_crimes_df['count'][i],
str(annual_crimes_df['count'][i]), color='red',
ha="center", va="bottom", fontsize=10)
plt.tight_layout()
plt.savefig('annual_crimes_trend.png')

# 2. Visualisasi Jenis Kejahatan (Crime Types)
plt.figure(figsize=(10, 6))
crime_type_crimes_sorted =
crime_type_crimes_df.sort_values(by='count', ascending=False)
sns.barplot(x='count', y='Crime',
data=crime_type_crimes_sorted, palette='viridis')
plt.title('Jumlah Kejahatan Berdasarkan Jenis')
plt.xlabel('Jumlah Kejahatan')
plt.ylabel('Jenis Kejahatan')

```



```

for i in range(len(crime_type_crimes_sorted)):
    plt.text(crime_type_crimes_sorted['count'].iloc[i] + 5, i,
             str(crime_type_crimes_sorted['count'].iloc[i]),
             color='blue', ha="left", va="center", fontsize=10)
plt.tight_layout()
plt.savefig('crime_type_distribution.png')

# 3. Visualisasi Kejahatan per Wilayah (Wilayah Crimes)
plt.figure(figsize=(10, 6))
wilayah_crimes_sorted =
wilayah_crimes_df.sort_values(by='count', ascending=False)
sns.barplot(x='count', y='Wilayah',
            data=wilayah_crimes_sorted, palette='coolwarm')
plt.title('Jumlah Kejahatan Berdasarkan Wilayah')
plt.xlabel('Jumlah Kejahatan')
plt.ylabel('Wilayah')
for i in range(len(wilayah_crimes_sorted)):
    plt.text(wilayah_crimes_sorted['count'].iloc[i] + 5, i,
             str(wilayah_crimes_sorted['count'].iloc[i]),
             color='green', ha="left", va="center", fontsize=10)
plt.tight_layout()
plt.savefig('wilayah_crimes_distribution.png')

```

Setelah tahap *data processing* selesai dilakukan, kode ini akan digunakan untuk memvisualisasikan hasil analisis data kejahatan yang telah diproses sebelumnya menggunakan Python dan beberapa library visualisasi seperti Matplotlib, Seaborn, dan Pandas.

Pertama-tama, kami perlu mengimpor beberapa library yang dibutuhkan. Library *pandas* digunakan untuk manipulasi data tabular dalam format *DataFrame*, sementara *matplotlib.pyplot* dan *seaborn* digunakan untuk membuat grafik visual yang mendalam dan estetik, sedangkan *pyspark.sql.Session* digunakan untuk membaca dan memproses data langsung dari HDFS. Dengan inisialisasi *SparkSession*, aplikasi diberi nama "*CrimeDataAnalysis*", dan fungsi *getOrCreate()* memastikan sesi *Spark* tersedia untuk pemrosesan data.

Setelah sesi *Spark* aktif, program menggunakan perintah *spark.read.csv()* untuk membaca file CSV yang berisi data kejahatan tahunan, berdasarkan jenis, dan per neighborhood dari HDFS. Data tersebut

kemudian diproses menggunakan Spark, dan dengan memanggil `toPandas()`, data dipindahkan ke Pandas `DataFrame` untuk memudahkan visualisasi menggunakan seaborn dan matplotlib

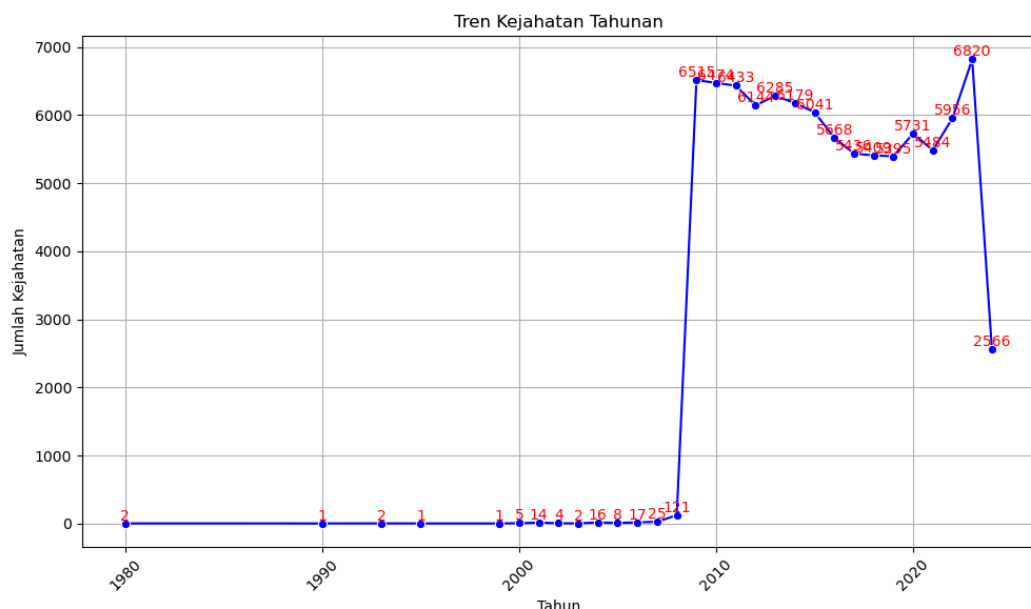
Visualisasi data kemudian dilakukan dalam tiga bagian utama, yaitu *line plot* untuk menunjukkan tren jumlah kejahatan per tahun, dengan nilai jumlah kejahatan ditampilkan pada setiap titik, *bar plot* untuk menunjukkan distribusi jumlah kejahatan berdasarkan jenis, dengan nilai di setiap bar ditampilkan, serta *bar plot* kedua untuk menganalisis distribusi kejahatan per neighborhood. Dan terakhir, setiap grafik kemudian akan disimpan sebagai file gambar dengan menggunakan `plt.savefig()`, agar hasil visualisasi dapat diakses dan dibagikan dalam format yang mudah digunakan.

Bab IV

Pembahasan

4.1. Tren Kejahatan Berdasarkan Tahun

Berdasarkan pengolahan data yang sudah dilakukan, diperoleh beberapa periode tahun yang mengalami perubahan tren kejahatan secara signifikan. Pada periode **2008** hingga **2009**, tercatat lonjakan drastis dalam jumlah kejahatan, dari 121 kasus pada tahun 2008 menjadi 6.515 pada 2009. Peningkatan tajam ini mungkin disebabkan oleh beberapa faktor penting. Salah satunya adalah perubahan dalam metode pencatatan, di mana sistem pelaporan digital mulai diadopsi. Dengan sistem yang lebih akurat, kejahatan yang sebelumnya tidak tercatat dapat dimasukkan ke dalam data, mencerminkan angka yang lebih realistis. Selain itu, peningkatan kesadaran masyarakat melalui kampanye pemerintah dan media mungkin juga berperan penting dalam peningkatan ini. Masyarakat yang lebih sadar akan pentingnya melaporkan kejahatan dapat menyebabkan jumlah laporan meningkat. Di sisi lain, krisis ekonomi global yang terjadi pada periode 2008-2009 turut memberikan tekanan ekonomi yang besar sehingga dapat mendorong peningkatan kejahatan properti dan kejahatan kecil.



Periode selanjutnya mencakup tahun **2010** hingga **2023**, di mana jumlah kejahatan tetap berada di angka tinggi dengan rata-rata sekitar 6.000 kejadian per tahun. Stabilitas ini mencerminkan beberapa kondisi penting dalam sistem

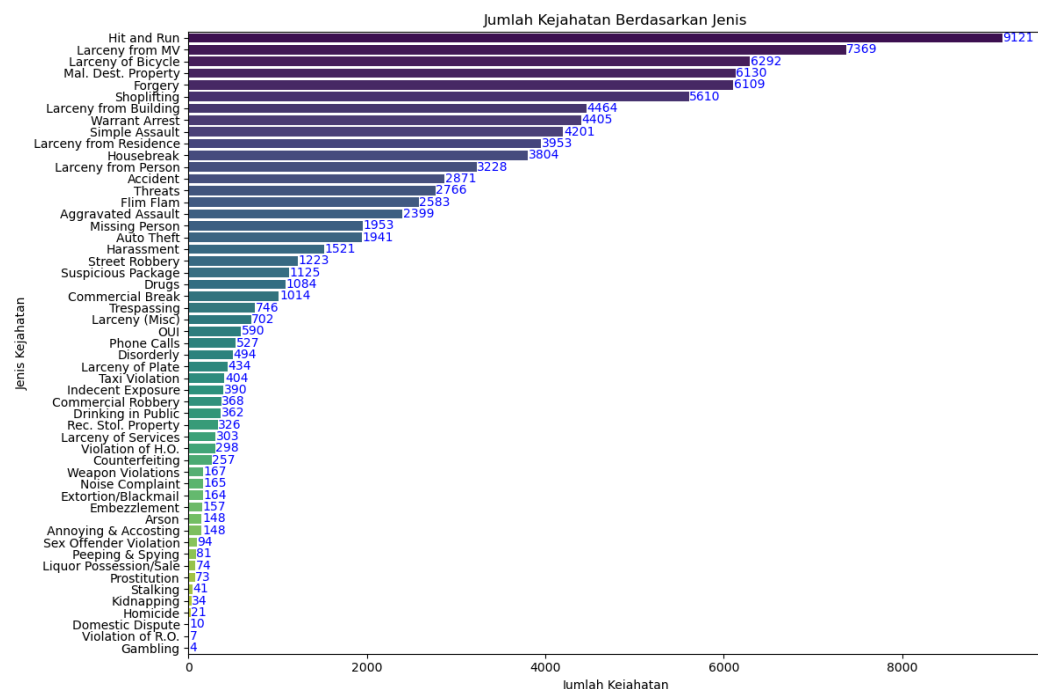
keamanan dan sosial di wilayah tersebut. Kemapanan sistem keamanan mungkin berhasil menahan lonjakan kejahatan lebih lanjut, namun belum cukup efektif untuk menurunkan angka secara signifikan. Pola kriminalitas tertentu, seperti pencurian, mungkin telah menjadi tantangan rutin yang sulit diberantas sepenuhnya. Selain itu, dinamika sosial-ekonomi yang stabil, termasuk pertumbuhan penduduk dan aktivitas ekonomi yang tidak banyak berubah, juga turut menjaga pola kejahatan tetap konsisten. Ini menunjukkan bahwa meskipun ada langkah-langkah penegakan hukum, sebenarnya tantangan keamanan struktural masih tetap ada.

Namun, pada tahun **2024** tercatat penurunan yang signifikan dalam aspek jumlah kejahatan, dengan penurunan angka yang lebih dari 50% atau menjadi 2.566 kasus dibandingkan dengan tahun sebelumnya. Penurunan ini dapat mencerminkan perubahan besar dalam sistem keamanan atau kondisi sosial. Intervensi teknologi modern, seperti penggunaan pengawasan berbasis kecerdasan buatan (AI), kamera pengenalan wajah, dan integrasi data penegakan hukum, dapat meningkatkan efektivitas dalam mencegah dan menangani kejahatan. Selain itu, kebijakan keamanan yang lebih ketat, seperti peningkatan patroli dan kerja sama antara polisi dengan masyarakat, juga dapat berkontribusi. Program edukasi masyarakat yang mendorong kesadaran akan pencegahan kejahatan mungkin turut memainkan peran penting. Namun, perlu diingat bahwa data tahun 2024 belum berakhir, sehingga angka ini masih dapat terus berubah seiring diperbarainya data di akhir tahun.

Analisis tren kejahatan dari 2008 hingga 2024 menunjukkan pola yang mencerminkan berbagai faktor sistemik, sosial, dan teknologi. Lonjakan pada 2009 menyoroti dampak perubahan pencatatan dan tekanan ekonomi, sedangkan stabilitas dari 2010 hingga 2023 menunjukkan tantangan struktural yang belum sepenuhnya teratasi. Penurunan besar pada 2024 memberikan harapan akan efektivitas kebijakan keamanan baru dan inovasi teknologi dalam mengatasi kejahatan. Keseluruhan tren ini menekankan pentingnya pendekatan multifaset, termasuk teknologi, kebijakan, dan edukasi masyarakat, untuk menciptakan lingkungan yang lebih aman dan kondusif bagi seluruh warga.

4.2. Jenis Kejahatan yang Paling Sering Terjadi

Berdasarkan pengolahan data yang telah dilakukan, jenis-jenis kejahatan yang dominan berhasil diidentifikasi, memberikan wawasan penting mengenai ancaman utama yang dihadapi wilayah tersebut. Analisis menunjukkan bahwa *Larceny* merupakan kejahatan yang paling sering terjadi, dengan total 24.859 kasus. Subkategori seperti *Larceny from MV* (7.369 kasus) dan *Larceny of Bicycle* (6.292 kasus) menonjol, menunjukkan perlunya peningkatan pengawasan di area parkir kendaraan dan penyediaan fasilitas parkir sepeda yang lebih aman. Selain itu, tingginya angka Hit and Run (9.121 kasus) mengindikasikan kurangnya penegakan hukum terhadap perilaku pengendara, serta perlunya pengelolaan lalu lintas yang lebih baik.



Kejahatan dengan dampak sosial yang signifikan, meskipun frekuensinya rendah, tetapi tetap menjadi perhatian. *Homicide* (21 kasus) dan *Kidnapping* (34 kasus) memiliki dampak psikologis dan sosial yang besar, memerlukan perhatian khusus dalam upaya pencegahan dan penanganannya. Selain itu, kategori kejahatan seperti Shoplifting (5.610 kasus) dan *Larceny from Building* (4.464 kasus) menunjukkan bahwa pelaku sering memanfaatkan celah keamanan di area komersial dan gedung dengan pengawasan minimum.

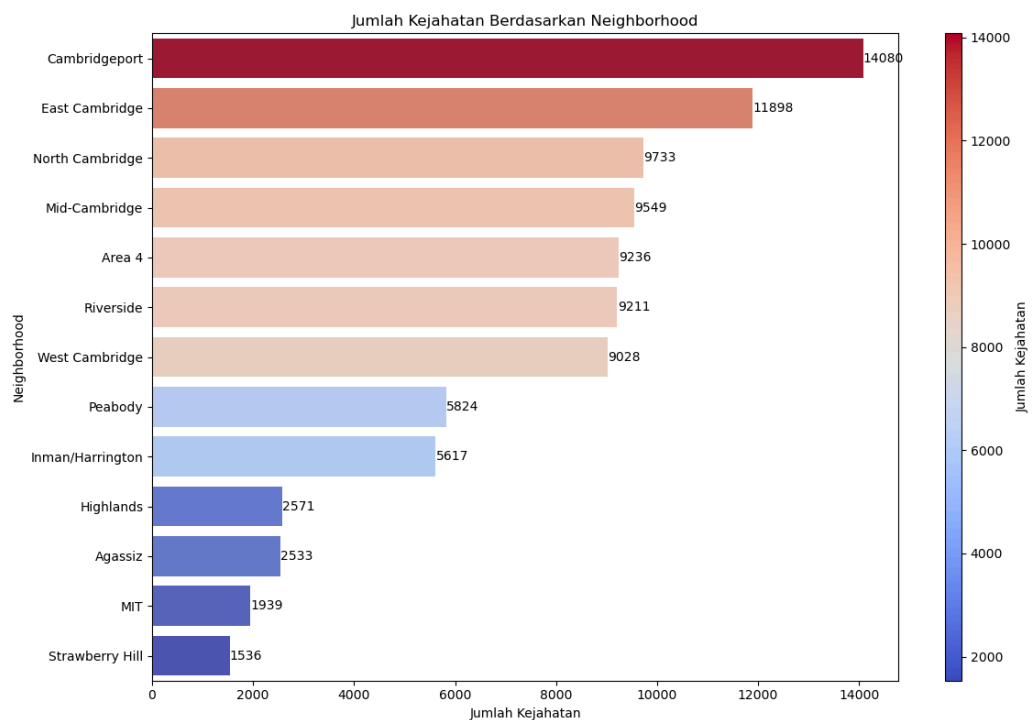
Tren khusus seperti *Auto Theft* (1.941 kasus) dan *Larceny of Plate* (434 kasus) menyiratkan perlunya penerapan sistem keamanan tambahan untuk

kendaraan pribadi. Masalah sosial seperti *Drinking in Public* (362 kasus) dan *Indecent Exposure* (390 kasus) mencerminkan kelemahan regulasi atau pengawasan di ruang publik yang ramai. Kejahatan berbasis konflik seperti *Aggravated Assault* (2.399 kasus) menekankan perlunya program mediasi masyarakat dan peningkatan patroli di area dengan tingkat konflik tinggi.

Hasil analisis ini menunjukkan bahwa dengan memahami pola jenis kejahatan yang sering terjadi, rekomendasi kebijakan yang lebih terfokus dapat diusulkan, termasuk peningkatan pengawasan, edukasi masyarakat, dan implementasi teknologi keamanan di area rawan. Analisis ini berhasil memenuhi tujuan proyek dalam memberikan wawasan berbasis data untuk mendukung kebijakan keamanan yang lebih efektif.

4.3. Distribusi Kejahatan Berdasarkan Wilayah

Berdasarkan pengolahan data, distribusi kejahatan di berbagai wilayah menunjukkan pola yang unik, dengan beberapa wilayah memiliki tingkat risiko kejahatan yang jauh lebih tinggi dibandingkan yang lain.



Berikut adalah analisis data kriminalitas yang kami peroleh berdasarkan wilayah:

4.3.1. Wilayah dengan Potensi Kejahatan Tinggi

- Cambridgeport (14.080 kasus)

Cambridgeport merupakan wilayah dengan tingkat kejahatan tertinggi. Faktor utama penyebabnya adalah kepadatan penduduk yang tinggi, yang sering kali meningkatkan peluang konflik atau kejahatan. Selain itu, banyaknya aktivitas komersial, seperti toko, restoran, dan pusat perbelanjaan, menarik keramaian yang menjadi sasaran empuk bagi pelaku kejahatan. Kemudahan akses transportasi juga memungkinkan mobilitas tinggi, baik bagi penduduk maupun pelaku kejahatan.

- East Cambridge (11.898 kasus)

Wilayah ini menempati posisi kedua dalam tingkat kejahatan. Kedekatan dengan area komersial dan transportasi utama menjadi faktor yang meningkatkan risiko kejahatan seperti pencurian dan kecelakaan. Kehadiran fasilitas publik, seperti stasiun kereta dan terminal bus, menciptakan keramaian yang sering kali menjadi titik rawan kejahatan.

4.3.2. Wilayah dengan Potensi Kejahatan Rendah

- Strawberry Hill (1.536 kasus)

Wilayah ini memiliki aktivitas yang rendah karena kepadatan penduduknya yang kecil serta minimnya kegiatan komersial. Dengan sedikitnya toko atau fasilitas umum besar, peluang terjadinya kejahatan di wilayah ini juga berkurang.

- MIT (1.939 kasus)

Sebagai area akademik, MIT menunjukkan tingkat kejahatan yang rendah. Hal ini dapat dikaitkan dengan pengawasan yang ketat oleh keamanan kampus dan keberadaan populasi sementara seperti mahasiswa, yang cenderung tidak menjadi target utama pelaku kejahatan.

4.3.3. Pola Unik Berdasarkan Wilayah

- Riverside (9.211 kasus) dan Mid-Cambridge (9.549 kasus)

Wilayah ini menunjukkan tingkat kejahatan yang relatif tinggi karena aktivitas sosial yang cukup intens. Kejahatan seperti vandalisme atau pencurian properti kecil sering terjadi di lingkungan dengan interaksi sosial yang tinggi.

- Peabody (5.824 kasus) dan Inman/Harrington (5.617 kasus)

Wilayah ini memiliki tingkat kejahatan sedang, yang mungkin disebabkan oleh kombinasi karakteristik residensial dan komersial. Aktivitas yang beragam ini menciptakan peluang kejahatan, meskipun tidak seintensif di wilayah seperti Cambridgeport.

4.4. Analisis Kombinasi

Hasil analisis menunjukkan hubungan erat antara jenis kejahatan tertentu dan distribusinya di berbagai wilayah. Berikut adalah hasil analisis berdasarkan kombinasi kejahatan dan wilayah:

4.4.1. Pencurian di Cambridgeport dan East Cambridge

Pencurian di Cambridgeport dan East Cambridge, termasuk Hit and Run dan berbagai kategori Larceny, didorong oleh tingginya jumlah fasilitas komersial seperti pusat perbelanjaan, restoran, dan toko yang menarik banyak orang. Kepadatan aktivitas ini menciptakan peluang lebih besar bagi pelaku kejahatan. Selain itu, area parkir terbuka dan stasiun transportasi umum yang minim pengawasan menjadi sasaran utama. Aksesibilitas transportasi yang tinggi juga memungkinkan pelaku kejahatan untuk bermobilitas dengan mudah, baik untuk melakukan kejahatan maupun melarikan diri.

Untuk mengurangi tingkat pencurian, diperlukan peningkatan patroli keamanan di area padat aktivitas selama jam sibuk. Sistem keamanan pintar, seperti kamera pengawas berbasis AI, dapat diinstal di lokasi strategis seperti area parkir, stasiun transportasi, dan persimpangan jalan. Edukasi masyarakat tentang tindakan preventif, seperti tidak meninggalkan barang berharga di kendaraan dan melaporkan aktivitas mencurigakan, juga penting untuk meningkatkan kesadaran akan keamanan.

4.4.2. Vandalisme di Riverside dan Mid-Cambridge

Vandalisme di Riverside dan Mid-Cambridge sering terjadi di area dengan aktivitas sosial yang tinggi, seperti taman kota, tempat hiburan, atau lokasi pertemuan komunitas. Aktivitas ini meningkatkan

interaksi sosial yang, jika tidak dikelola dengan baik, dapat memicu konflik atau perilaku destruktif. Kurangnya pencahayaan atau pengawasan di area publik tertentu juga memberikan peluang lebih besar bagi pelaku vandalisme untuk beraksi tanpa terdeteksi.

Untuk mengurangi tingkat vandalisme, diperlukan peningkatan infrastruktur, seperti menambahkan pencahayaan di tempat berkumpul atau area publik lainnya, untuk menciptakan lingkungan yang lebih aman. Selain itu, kampanye sosial yang mengajak masyarakat berkontribusi dalam menjaga kebersihan dan keamanan lingkungan dapat meningkatkan rasa memiliki terhadap area tersebut. Memberlakukan sanksi hukum yang lebih berat bagi pelaku vandalisme juga diperlukan untuk menciptakan efek jera dan mencegah kejahatan serupa di masa depan.

4.4.3. Kejahatan Berat yang Tersebar Secara Merata

Kejahatan berat seperti homicide dan kidnapping tidak terkonsentrasi di satu wilayah tertentu, melainkan tersebar merata. Kejahatan ini lebih bergantung pada faktor situasional, seperti konflik pribadi, kesempatan, atau perencanaan terisolasi, dibandingkan dengan karakteristik wilayah tertentu. Minimnya pola yang konsisten membuat kejahatan ini sulit untuk diprediksi atau dicegah menggunakan pendekatan tradisional.

Untuk menangani kejahatan berat sendiri, diperlukan penelitian mendalam untuk menganalisis pola berdasarkan waktu kejadian, hubungan korban dan pelaku, serta faktor pemicu lainnya. Integrasi data antar-lembaga penegakan hukum juga diperlukan untuk mendeteksi pola tersembunyi dan mencegah kasus serupa. Selain itu, peningkatan respons terhadap laporan aktivitas mencurigakan, dengan memanfaatkan sistem pelaporan yang terpadu, dapat membantu dalam mencegah eskalasi kejahatan berat.

Distribusi kejahatan menunjukkan bahwa wilayah dengan aktivitas komersial dan sosial tinggi, seperti Cambridgeport dan East Cambridge, menghadapi risiko yang lebih besar. Sebaliknya, kejahatan berat seperti Homicide dan Kidnapping tidak bergantung pada lokasi tertentu, melainkan bergantung

pada faktor situasional. Dengan strategi mitigasi yang tepat, seperti peningkatan patroli keamanan, penggunaan teknologi canggih, dan kampanye sosial, lingkungan yang lebih aman dapat diciptakan.

4.5. Tren Sosial dan Kriminalitas

Tingkat kriminalitas setiap tahun maupun pada setiap wilayah dapat dipengaruhi oleh berbagai faktor, mulai dari teknologi, ekonomi, hingga perubahan sosial. Seperti pada tahun 2024, penurunan signifikan dalam angka kejahatan sebagian besar dapat dikaitkan dengan implementasi teknologi keamanan yang canggih. Teknologi seperti pengawasan berbasis kecerdasan buatan (AI), kamera pengenalan wajah, dan integrasi big data telah meningkatkan efisiensi dalam mendeteksi dan mencegah kejahatan. Kehadiran teknologi ini tidak hanya memberikan pengawasan lebih ketat, tetapi juga menciptakan efek jera dengan mempersempit peluang pelaku untuk beraksi tanpa terdeteksi. Penggunaan sistem otomatis ini menjadikan keamanan lebih responsif dan proaktif, menciptakan lingkungan yang lebih aman. Jika strategi ini diadopsi secara lebih luas, terutama di wilayah dengan risiko kejahatan tinggi, potensi penurunan kejahatan di masa depan semakin besar.

Sebaliknya, lonjakan tajam angka kejahatan tercatat pada tahun 2009 menunjukkan bahwa adanya dampak yang kuat dari krisis ekonomi global pada tingkat kriminalitas. Di mana ketika ekonomi dunia menghadapi ketidakstabilan, tekanan finansial yang meningkat seringkali mendorong peningkatan kejahatan, terutama yang berkaitan dengan properti seperti pencurian dan vandalisme. Krisis tersebut menciptakan situasi di mana individu merasa terdorong untuk melakukan aktivitas ilegal sebagai cara memenuhi kebutuhan mendesak. Keadaan ini menunjukkan betapa pentingnya menjaga stabilitas ekonomi untuk mengurangi insentif terhadap tindakan kriminal. Untuk menghadapi krisis serupa di masa depan, program perlindungan sosial dan pemberdayaan ekonomi yang komprehensif menjadi kunci. Dengan memberikan dukungan kepada masyarakat yang rentan, risiko peningkatan kejahatan dapat diminimalisir secara signifikan.

Data yang tersedia menunjukkan bahwa pada periode sebelum tahun 2009, angka kejahatan yang tercatat dapat dikatakan sangat rendah. Namun, data ini tidak selalu mencerminkan situasi sebenarnya, melainkan kurangnya kesadaran atau kepercayaan masyarakat terhadap sistem hukum. Banyak insiden

yang tidak dilaporkan karena berbagai alasan, termasuk ketakutan akan represalias atau skeptisisme terhadap penegakan hukum. Dengan diperkenalkannya sistem pelaporan digital dan kampanye edukasi, pelaporan menjadi lebih mudah dan transparan, sehingga menciptakan data yang lebih akurat.

Tren sosial juga memainkan peran besar dalam mempengaruhi pola kriminalitas. Perubahan demografi, seperti migrasi penduduk dari daerah padat menuju lingkungan dengan tingkat keamanan lebih tinggi, dapat menyebabkan penurunan angka kejahatan di beberapa wilayah. Sebaliknya, daerah dengan kepadatan penduduk tinggi sering kali menghadapi tantangan keamanan yang lebih besar. Kampanye sosial yang mendorong kesadaran tentang dampak negatif kejahatan juga berkontribusi dalam menciptakan lingkungan yang lebih aman. Program-program ini menanamkan rasa tanggung jawab kolektif di masyarakat, mendorong partisipasi aktif dalam menjaga keamanan lingkungan mereka sendiri. Dengan komunitas yang lebih sadar akan pentingnya keamanan, potensi untuk mencegah tindakan kriminal meningkat.

Kerja sama antara komunitas lokal dan penegak hukum juga menjadi kunci penting dalam menciptakan lingkungan yang aman dan kondusif. Ketika komunitas merasa terhubung dan memiliki kepercayaan terhadap pihak berwenang, kolaborasi ini menghasilkan efektivitas yang lebih tinggi dalam pencegahan kejahatan. Inisiatif berbasis komunitas, seperti patroli warga dan forum keamanan, membantu memperkuat hubungan antara masyarakat dan aparat penegak hukum. Kolaborasi ini tidak hanya meningkatkan rasa aman, tetapi juga menciptakan kepercayaan yang lebih kuat terhadap sistem hukum. Dengan memperkuat kerja sama ini, potensi kejahatan dapat ditekan secara signifikan, sehingga akan menciptakan lingkungan yang lebih harmonis dan aman bagi semua pihak.

4.6. Solusi dan Rekomendasi

Berdasarkan hasil analisis yang telah diperoleh, kami menemukan beberapa langkah strategis yang dapat diterapkan untuk dapat menangani tingginya angka kriminalitas yang terjadi di beberapa wilayah. Berikut langkah-langkah strategis tersebut:

- **Pencegahan Pencurian**

Untuk mencegah kejahatan oportunistik, langkah pertama yang dapat diambil adalah meningkatkan pencahayaan di area publik, seperti jalan, tempat parkir, dan gang sempit, dengan lampu yang terang dan tahan lama. Pencahayaan yang memadai dapat mengurangi peluang kejahatan dengan membuat pelaku merasa lebih terlihat dan terdeteksi. Selain itu, pemasangan kamera pengawas pintar dengan teknologi AI, seperti pengenalan wajah dan analisis pola gerak, dapat memperkuat deteksi aktivitas mencurigakan di area yang berisiko tinggi. Kamera ini dapat memberikan pemantauan 24/7 dan bekerja sama dengan pusat kontrol keamanan untuk memberikan respons cepat. Selain itu, penting untuk mengedukasi masyarakat mengenai keamanan pribadi, seperti tidak meninggalkan barang berharga di kendaraan dan menggunakan kunci tambahan pada sepeda atau kendaraan pribadi, guna mengurangi potensi kejahatan yang terjadi akibat kelalaian. Pendekatan ini secara kolektif akan meningkatkan keamanan dan mengurangi risiko kejahatan di ruang publik.

- **Kebijakan untuk Jenis Kejahatan Khusus atau Tertentu**

Memberlakukan sanksi tegas terhadap pelanggaran seperti Hit and Run dan vandalisme dapat berfungsi sebagai pencegah yang efektif, karena memberikan pesan yang jelas bahwa pelanggaran terhadap hukum akan berakibat serius. Dana yang diperoleh dari denda tersebut dapat dialokasikan untuk memperbaiki kerusakan yang disebabkan oleh pelanggaran, serta untuk mendukung program keamanan komunitas yang lebih luas. Selain itu, edukasi masyarakat sangat penting dalam mencegah kejahatan, terutama shoplifting, yang banyak terjadi di kalangan remaja. Kampanye dan workshop yang meningkatkan kesadaran tentang dampak negatif dari perilaku ini, serta pendekatan berbasis sekolah dan komunitas, dapat mengurangi kecenderungan melakukan kejahatan tersebut sejak dini. Kolaborasi antara pihak berwenang dan bisnis juga sangat diperlukan untuk memperketat sistem keamanan di toko dan pusat perbelanjaan. Dengan melibatkan pemilik usaha dalam menerapkan teknologi anti-pencurian, seperti detektor barang curian dan pengawasan tambahan, risiko terjadinya pencurian dapat diminimalkan, menciptakan lingkungan yang lebih aman bagi konsumen dan pemilik usaha.

- **Pendekatan Spesifik Berdasarkan Wilayah**

Untuk area seperti Cambridgeport dan East Cambridge, pendekatan strategis yang fokus pada peningkatan patroli keamanan dapat sangat membantu. Menambah jumlah patroli polisi, terutama pada malam hari, akan memperkuat kehadiran otoritas yang terlihat dan memberikan rasa aman bagi warga. Selain itu, menciptakan zona aman dengan pengawasan 24 jam di tempat-tempat publik yang sering menjadi lokasi kejadian kejahatan dapat meminimalisir potensi ancaman. Pengawasan ini dapat dilakukan melalui kamera pengawas atau patroli rutin oleh petugas. Selain upaya yang dilakukan oleh pihak berwenang, kemitraan masyarakat juga memainkan peran penting dalam meningkatkan keamanan. Menggalang dukungan warga untuk berpartisipasi dalam program Neighborhood Watch, yang memungkinkan laporan real-time kepada pihak berwenang, dapat mempercepat respons terhadap kejadian kriminal dan mengurangi tingkat kejahatan secara signifikan. Program ini tidak hanya meningkatkan kewaspadaan, tetapi juga dapat memperkuat hubungan antara komunitas dan aparat keamanan, menciptakan rasa tanggung jawab kolektif yang penting dalam menjaga lingkungan tetap aman.

Pendekatan berbasis komunitas di Riverside dan Mid-Cambridge dapat mengurangi potensi konflik sosial dan mencegah kejahatan dengan mengadakan program positif seperti acara budaya, olahraga, dan lokakarya seni untuk mempererat hubungan antarwarga. Pendirian pusat mediasi untuk menyelesaikan perselisihan kecil juga dapat mencegah eskalasi masalah. Selain itu, sosialisasi mengenai dampak vandalisme melalui kampanye berbasis sekolah dan komunitas, yang menyoroti kerugian finansial dan sosial, dapat mengubah perilaku negatif dan meningkatkan kesadaran warga tentang pentingnya menjaga fasilitas umum. Semua inisiatif yang melalui kolaborasi komunitas dan pihak berwenang ini diharapkan dapat menciptakan lingkungan yang lebih aman dan harmonis.

4.7. Kemungkinan Analisis Lanjutan

Setelah analisis tingkat kriminalitas ini selesai dilakukan, kami menemukan beberapa kemungkinan analisis lanjutan yang dapat dirancang kedepannya. Berikut beberapa kemungkinan analisis lanjutan yang kami rekomendasikan:

4.7.1. Clustering Wilayah Berdasarkan Jenis Kejahatan

Penggunaan algoritma clustering seperti K-Means atau DBSCAN dapat mengelompokkan wilayah berdasarkan pola kejahatan dominan, seperti pencurian, vandalisme, atau kejahatan berat. Dengan hasil clustering ini, kebijakan keamanan dapat lebih terfokus dan spesifik sesuai dengan karakteristik setiap kelompok. Misalnya, area dengan tingkat pencurian tinggi dapat difokuskan pada peningkatan patroli keamanan, sementara wilayah dengan konflik sosial tinggi dapat diarahkan pada program komunitas dan mediasi. Visualisasi hasil clustering melalui peta tematik atau heatmap memberikan gambaran yang jelas tentang distribusi kejahatan, sehingga pembuat kebijakan dapat merumuskan langkah-langkah yang terukur dan tepat sasaran untuk setiap area yang berbeda.

4.7.2. Prediksi Tren dengan Time Series Analysis

Model seperti ARIMA atau LSTM sangat efektif dalam menganalisis tren kejahatan historis dan memprediksi pola di masa depan, termasuk mengidentifikasi musim atau periode tertentu dengan risiko tinggi. Dengan prediksi ini, pihak berwenang dapat mengalokasikan sumber daya dengan lebih efisien, misalnya dengan menempatkan petugas patroli tambahan di waktu-waktu tertentu atau memperkuat operasi keamanan di masa rawan. Integrasi prediksi tren dengan sistem pengawasan real-time juga memungkinkan pembentukan sistem peringatan dini yang mampu mendeteksi potensi eskalasi kejahatan secara cepat, sehingga tindakan preventif dapat diambil sebelum situasi memburuk. Pendekatan ini dapat menciptakan respons keamanan yang lebih proaktif dan strategis, meningkatkan perlindungan masyarakat secara keseluruhan.

4.7.3. Analisis Korelasi Sosio-Ekonomi

Mengintegrasikan data sosio-ekonomi, seperti tingkat pendapatan, pengangguran, pendidikan, dan kepadatan penduduk, dengan data kriminal dapat memberikan wawasan mendalam tentang hubungan antara kondisi sosial dan jenis kejahatan tertentu. Misalnya, wilayah dengan tingkat pendapatan rendah sering kali lebih rentan terhadap kejahatan properti, sementara wilayah dengan kepadatan tinggi berpotensi memicu konflik antar-warga. Analisis ini membantu mengidentifikasi faktor-faktor mendasar yang mendorong aktivitas kriminal, sehingga solusi yang diusulkan dapat lebih tepat sasaran. Program berbasis masyarakat, seperti pelatihan kerja untuk mengurangi pengangguran, beasiswa untuk meningkatkan akses pendidikan, atau peningkatan fasilitas umum, dapat dirancang khusus untuk mengatasi masalah ini di area yang membutuhkan. Pendekatan ini tidak hanya menangani gejala kriminalitas, tetapi juga berupaya untuk mengatasi akar penyebab, serta menciptakan lingkungan yang lebih aman dan stabil dalam jangka panjang.

Bab V

Kesimpulan

Dengan memanfaatkan tools big data, analisis tingkat kriminalitas di Cambridge dapat dilakukan dengan lebih efektif, terutama dalam mengelola dan menganalisis data yang besar dan kompleks. Hadoop berperan penting dalam menyimpan dan mengelola volume data yang sangat besar, seperti data laporan kejahatan dari berbagai sumber. Dengan menggunakan Spark, data tersebut diproses dengan cepat untuk mengekstrak informasi yang relevan mengenai tren kejahatan berdasarkan lokasi dan waktu. Misalnya, data dari tahun 2010 hingga 2023 menunjukkan fluktuasi jumlah kejahatan yang cukup signifikan, dengan lonjakan drastis pada tahun 2009, yang mungkin disebabkan oleh perubahan metodologi pelaporan atau dampak krisis ekonomi global. Melalui Spark, identifikasi pola ini dapat dilakukan dengan lebih cepat dan akurat, memungkinkan analisis lebih mendalam tentang faktor-faktor yang mempengaruhi lonjakan tersebut.

Hasil analisis menunjukkan bahwa sejumlah area di Cambridge, seperti Cambridgeport dan East Cambridge, memiliki tingkat kejahatan yang lebih tinggi, terutama dalam kategori pencurian dan vandalisme. Visualisasi data menggunakan Matplotlib dalam bentuk heatmap memudahkan pemahaman tentang distribusi kejahatan ini, mengungkapkan wilayah yang memerlukan perhatian lebih dari pihak berwenang. Selain itu, dengan menganalisis faktor waktu, ditemukan pola musiman dalam kejahatan tertentu, seperti peningkatan kasus pencurian pada akhir pekan atau selama musim liburan. Data ini dapat digunakan oleh pihak berwenang untuk merencanakan alokasi sumber daya yang lebih efisien, seperti penempatan patroli yang lebih intensif di wilayah-wilayah dengan risiko tinggi.

Berdasarkan hasil analisis ini, rekomendasi strategis untuk pencegahan kejahatan dapat difokuskan pada peningkatan pengawasan di area dengan tingkat kejahatan tinggi serta memperkuat kolaborasi antara komunitas dan aparat penegak hukum. Selain itu, kebijakan yang berbasis pada data yang telah dianalisis ini dapat membantu mengurangi kejahatan dengan memfokuskan sumber daya pada waktu-waktu rawan dan lokasi yang membutuhkan perhatian khusus. Dengan integrasi Hadoop, Spark, dan Matplotlib dalam analisis data kriminal, diharapkan Cambridge dapat lebih siap menghadapi tantangan dalam menjaga keamanan masyarakat, serta merancang kebijakan yang lebih tepat sasaran dan berbasis bukti.

Referensi

- [1] N. Ruddhan, "Crime Analysis & Feature Engineering," Kaggle, 2021. [Online]. Available: <https://www.kaggle.com/code/naniruddhan/crimeanalysis-featureengineering/input>. [Accessed: Nov. 14, 2024].
- [2] "Crime Trends," Fiveable, [Online]. Available: <https://fiveable.me/key-terms/comparative-criminal-justice-systems/crime-trends>. [Accessed: Nov. 14, 2024].
- [3] "Hadoop," IBM, [Online]. Available: <https://www.ibm.com/id-id/topics/hadoop>. [Accessed: Nov. 14, 2024].
- [4] "Hadoop vs Spark," IBM, [Online]. Available: <https://www.ibm.com/think/insights/hadoop-vs-spark>. [Accessed: Nov. 14, 2024].
- [5] "Introduction to Matplotlib in Python," GeeksforGeeks, [Online]. Available: <https://www.geeksforgeeks.org/python-introduction-matplotlib/>. [Accessed: Nov. 14, 2024].