

LAPORAN EKSPERIMEN

Pengantar Data Science

Explorasi Dataset Forest Fires

forestfire.csv

Disusun Oleh

William Surya Wijaya – 6182001004
Kae Zarren Luvoma H – 6182001004
Ester Oka Nugraheni – 6182001042

Tanggal

Sabtu, 26 November 2022

Jurusan Informatika

UNIVERSITAS KATOLIK PARAHYANGAN

I. Tujuan Eksperimen

Tujuan dari eksperimen dari dataset **forestfires.csv** adalah untuk mengetahui dan mengenali data beserta hubungan antar atribut untuk mengambil kesimpulan mengenai faktor faktor yang berhubungan dengan luas kebakaran hutan dan intensitas kebakaran hutan.

II. Deskripsi Dataset, Penyiapan Data, dan Hipotesis

Dataset **forestfires.csv** memiliki dimensi (517, 13)

Berikut adalah deskripsi dataset berupa nama atribut, keterangan beserta nilai minimal dan maksimal dari masing masing atribut.

No	Atribut	Keterangan	Min	Max
1.	X	Koordinat x dari lokasi kebakaran hutan di dalam lingkup area Montesinho Park.	1	9
2.	Y	Koordinat y dari lokasi kebakaran hutan di dalam lingkup area Montesinho Park.	2	9
3.	Month	Bulan terjadinya kebakaran hutan (January – December) dengan format 3 first letters.	Jan	Dec
4.	Day	Hari terjadinya kebakaran hutan (Monday – Sunday) dengan format 3 first letters.	Mon	Sun
5.	FFMC	Fine Fuel Moisture Code - Index tingkat potensi terjadinya kebakaran yang diukur dari kelembaban serasah dan bahan halus mudah terbakar di sekitar.	18.7	96.2
6.	DMC	Duff Moisture Code - Index tingkat potensi terjadinya kebakaran yang diukur dari kelembaban rata rata lapisan tanah organik dan bahan berukuran sedang yang mudah terbakar di sekitar.	1.1	291.3
7.	DC	Drought Code – Index tingkat potensi terjadinya kebakaran yang diukur dari kelembaban lapisan tanah organik padat dan bahan berukuran besar (bongkahan) yang mudah terbakar di sekitar dan potensi terjadinya kabut asap.	7.9	860.6
8.	ISI	Initial Spread Index - Index perkiraan penyebaran api yang diukur dari FFMC dan pergerakan angin.	0.0	56.1
9.	Temp	Temperatur sekitar dalam satuan derajat celcius.	2.2	33.3
10.	RH	Kelembaban sekitar dalam satuan %.	15	100
11.	Wind	Kecepatan angin dalam satuan km/h.	0.4	9.4
12.	Rain	Kepadatan hujan dengan satuan mm/m2.	0.0	64
13.	Area	Luas area kebakaran hutan dengan satuan ha.	0.0	1090.84

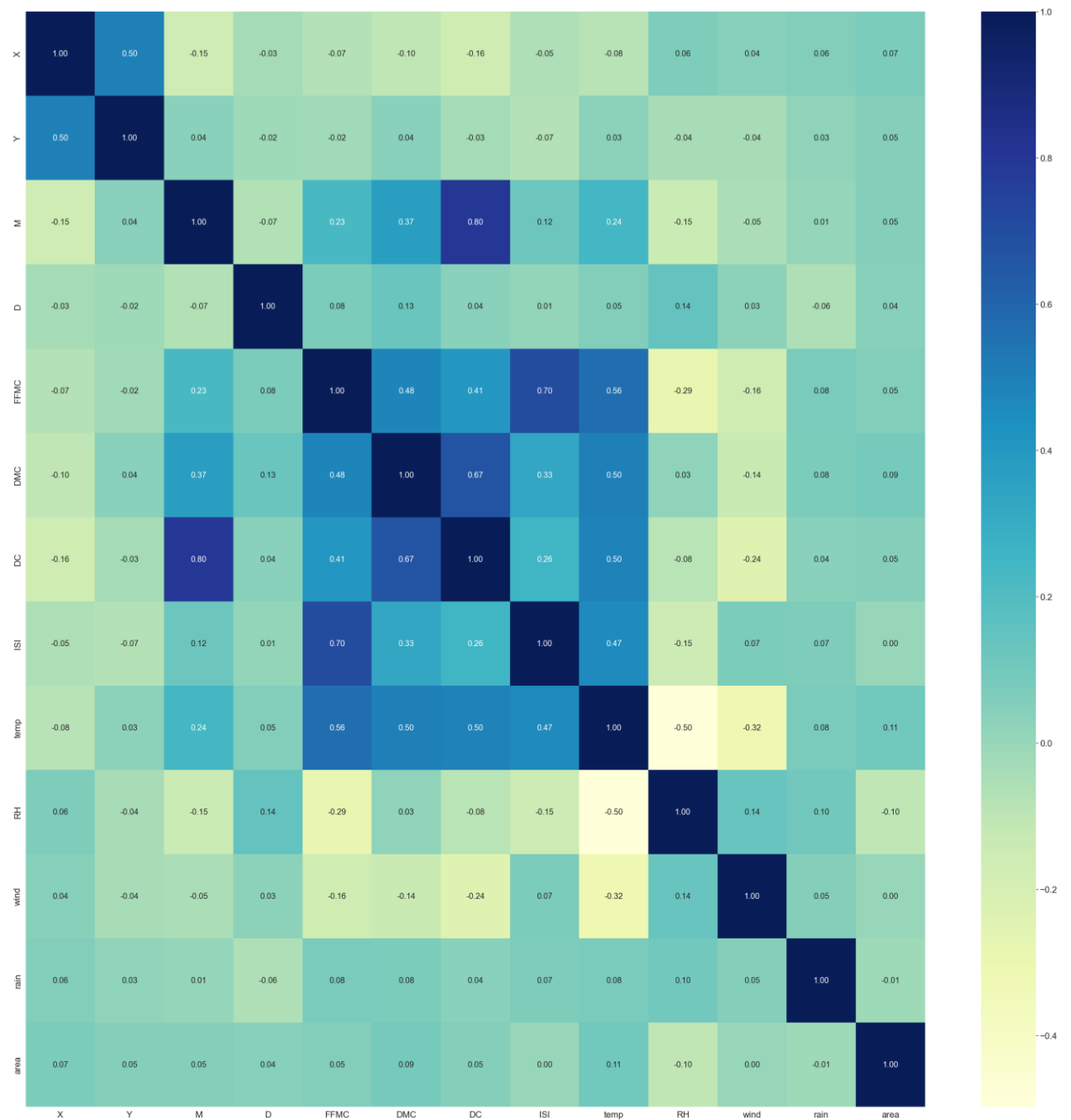
Dataset **forestfires.csv** tidak memiliki nilai NaN.

Hipotesis dan Hipotesa

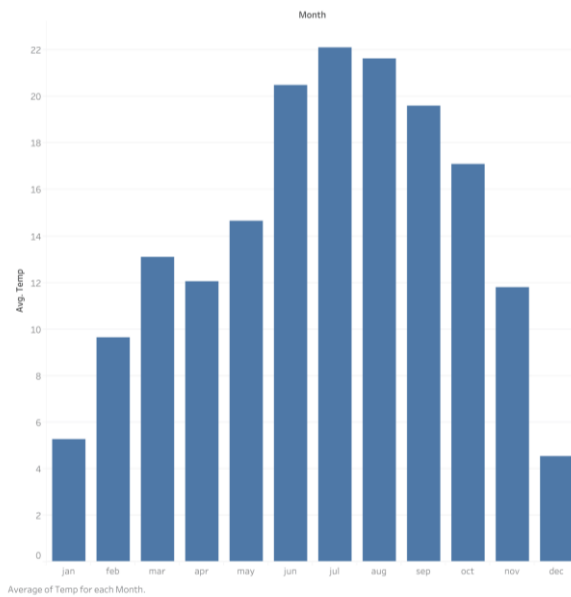
1. Apa faktor yang berpengaruh secara signifikan terhadap atribut target (area)?
2. Pada bulan mana sajakah yang memiliki tingkat intensitas kebakaran hutan tinggi?
3. Pada bulan mana sajakah yang memiliki luas area kebakaran hutan tinggi?
4. Pada bulan mana sajakah yang memiliki tingkat temperatur tinggi?

III. Penelitian

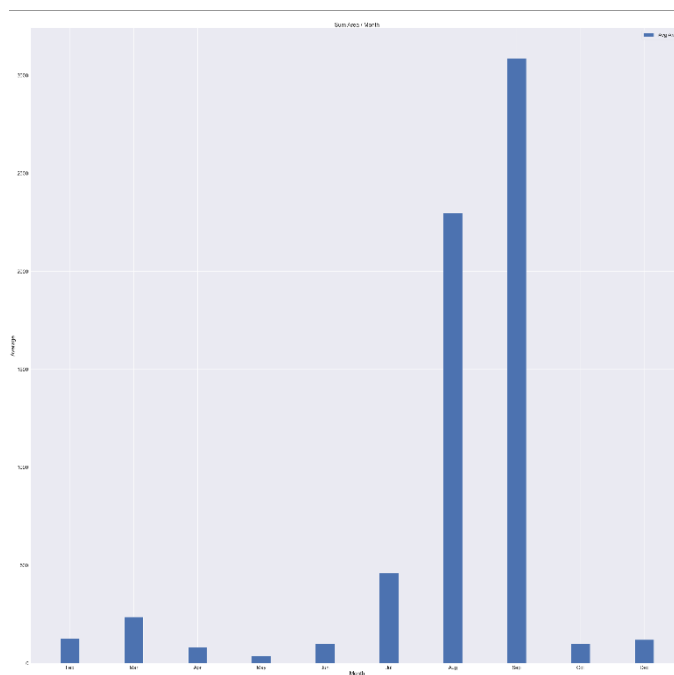
Mengetahui atribut atribut yang memiliki pengaruh signifikan baik hubungan positif dan hubungan negatif terhadap atribut target (area) dengan menggunakan metode Heat-Map.



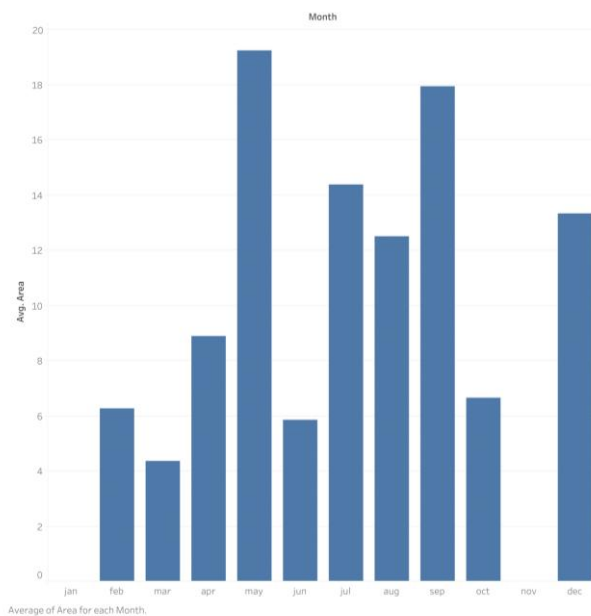
Visualisasi hubungan atribut atribut prediktor dengan atribut target tidak menunjukkan hubungan korelasi yang kuat. Berikut adalah beberapa atribut dengan nilai korelasi tertinggi.



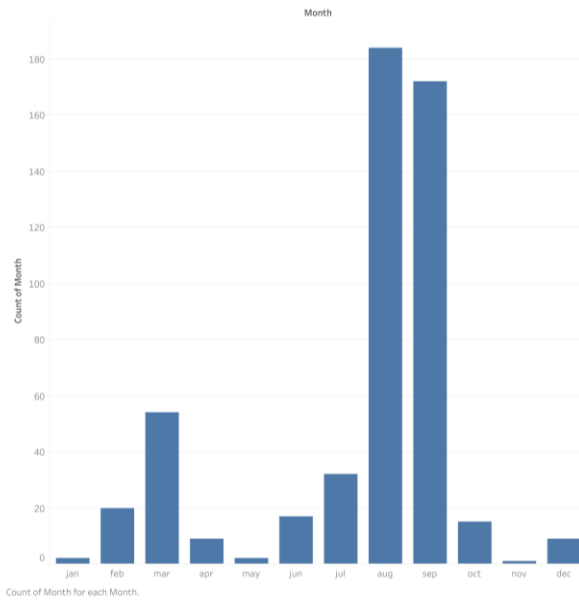
Visualisasi perbandingan dari rata-rata suhu temperature dan bulan yang dimana sering terjadi kebakaran pada bulan Juli (22.11 °C) dan Agustus (21.63 °C) dikarenakan oleh suhu yang tinggi.



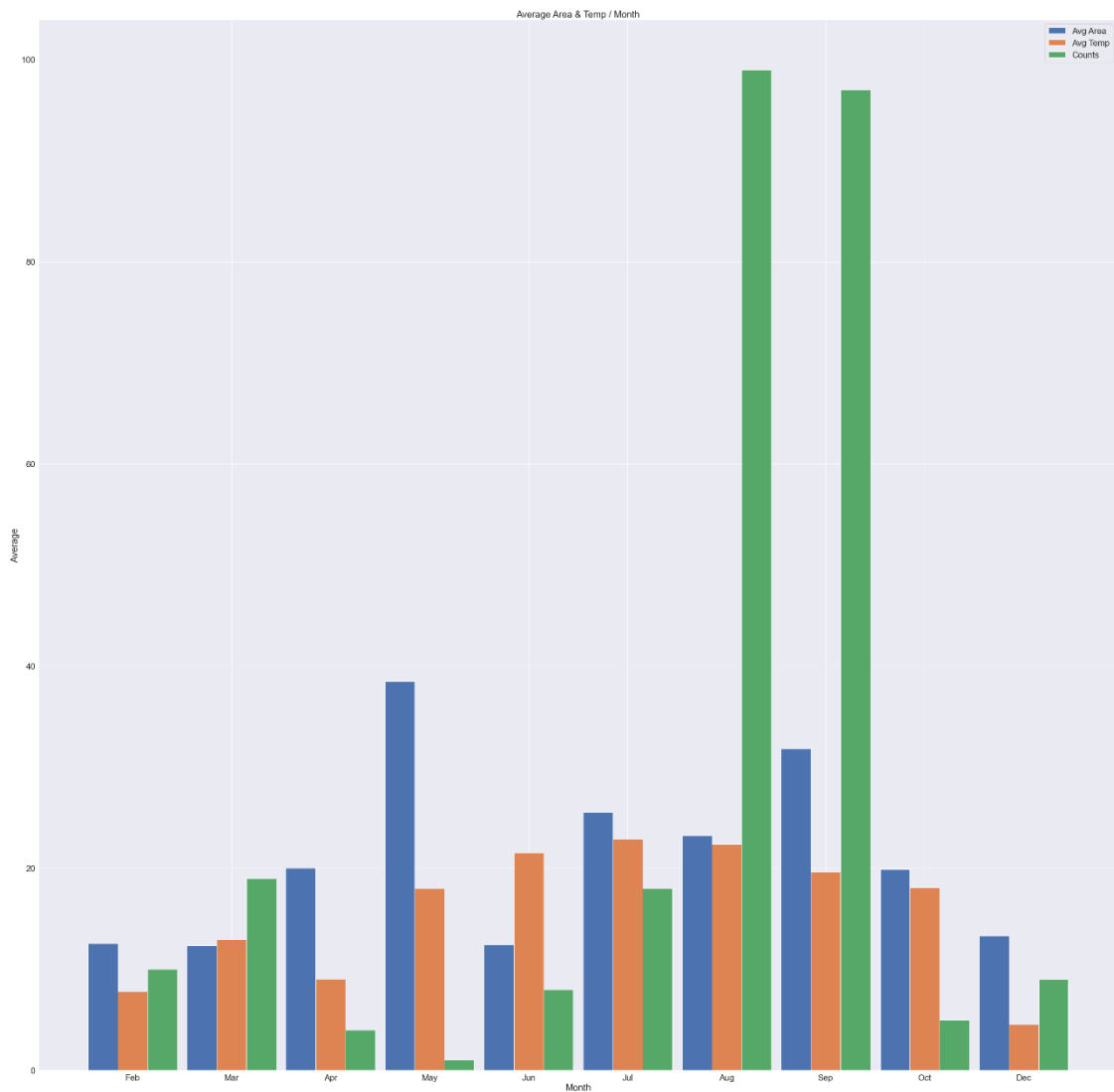
Visualisasi perbandingan dari total luas area yang terbakar per bulan, dimana total luas area terbakar yang tertinggi ada pada bulan September dan Agustus.



Visualisasi perbandingan dari rata-rata luas area yang terbakar dan bulan (month) yang dimana rata-rata luas area kebakaran yang paling besar adalah pada bulan Mei.



Visualisasi bulan (month) yang sering terjadi kebakaran hutan adalah pada bulan Agustus (184 kebakaran) & September (172 kebakaran)



Visualisasi perbandingan antara rata-rata area yang terbakar (bar biru), rata-rata suhu (bar oren) dan total kejadian (bar hijau). Jika dilihat dari visualisasi di atas, banyak kejadian kebakaran hutan terjadi pada bulan Agustus dan September. Area yang sering terjadi kebakaran adalah pada bulan Mei dan diikuti pada bulan September. Sedangkan rata-rata suhu tertinggi ada pada bulan Juli dan diikuti oleh bulan Agustus.

IV. Kesimpulan

Dengan menggunakan visualisasi Bar Chart dan Heatmap, kita dapat menganalisis secara mendalam pada kasus dataset forestfires.csv. Dalam laporan ini, kami telah mengetahui beberapa kejadian kebakaran hutan dari suhu rata rata perbulan, total area yang terbakar, dan juga pada bulan beberapa banyak terjadi kebakaran hutan per bulan.

Lampiran

<https://github.com/shrk-sh-ioai/pds-banking-dt/tree/main/tugas-akhir/assets>