**LAPORAN EKSPERIMEN**

Pengantar Data Science

**Explorasi Dataset Bike**

day.csv - hour.csv

**Disusun Oleh**

William Surya Wijaya | 6182001004

Kae Zarren Luvoma Halim | 6182001005

Ester Oka Nugraheni | 6182001042

**Tanggal**

Sabtu,26 November 2022

**Jurusan Informatika**

**UNIVERSITAS KATOLIK PARAHYANGAN**

1. **Tujuan Eksperimen**

Tujuan dari eksperimen dari data ***day.csv*** *dan* ***hour.csv***adalah untuk mengetahui dan mengenali data berserta hubungan antar atribut untuk mengambil kesimpulan mengenai faktor yang berhubungan total penyewaan sepeda.

1. **Deskripsi Dataset, Penyiapan Data, dan Hipotesis**

Atribut data yang ada pada dataset **day** dan **hour** adalah sebagai berikut.

instant

dteday

season

yr

mnth

hr

holiday

weekday

workingday

weathersit

temp

atemp

hum

windspeed

casual

registered

cnt

Dataset **day** dan **hour** tidak memilihi nilai NA, null, dan error di setiap atribut yang ada.

Hipotesis dan Hipotesa

1. Apa faktor faktor yang berpengaruh secara signifikan terhadap atribut target ?
2. Penelitian

Mengetahui variable variable yang memiliki pengaruh signifikan baik hubungan positif dan negatif terhadap atribut target dengan metode Chi-Square dan Heat-Map.

Metode Chi-Square dan Heat-Map memiliki hasil yang berbeda yaitu, Chi-Square mendapatkan 4 variable tertinggi yaitu year, month, holiday, dan weekday, sementara Heat-Map mendapatkan 4 variable tertinggi yaitu temp, year, season, dan month.

Berikut adalah hasil dari Chi-Square dan Heat-Map.

**Chi-Square**

#metode selection (1) : chi-square

k = 4;

selector = SelectKBest(*score\_func* = chi2, *k*=4);

selector.fit(day\_features\_np, day\_labels\_np);

cols = selector.get\_support(*indices*=True);

day\_features = day\_features.iloc[:,cols];

day\_features\_np = np.array(day\_features.values);

print (str(k)+" Kolom dengan nilai Chi Square terbaik adalah ")

print(day\_features.columns);

Index([‘yr’,'mnth', 'holiday', 'weekday'], dtype='object')

**Heat-Map**

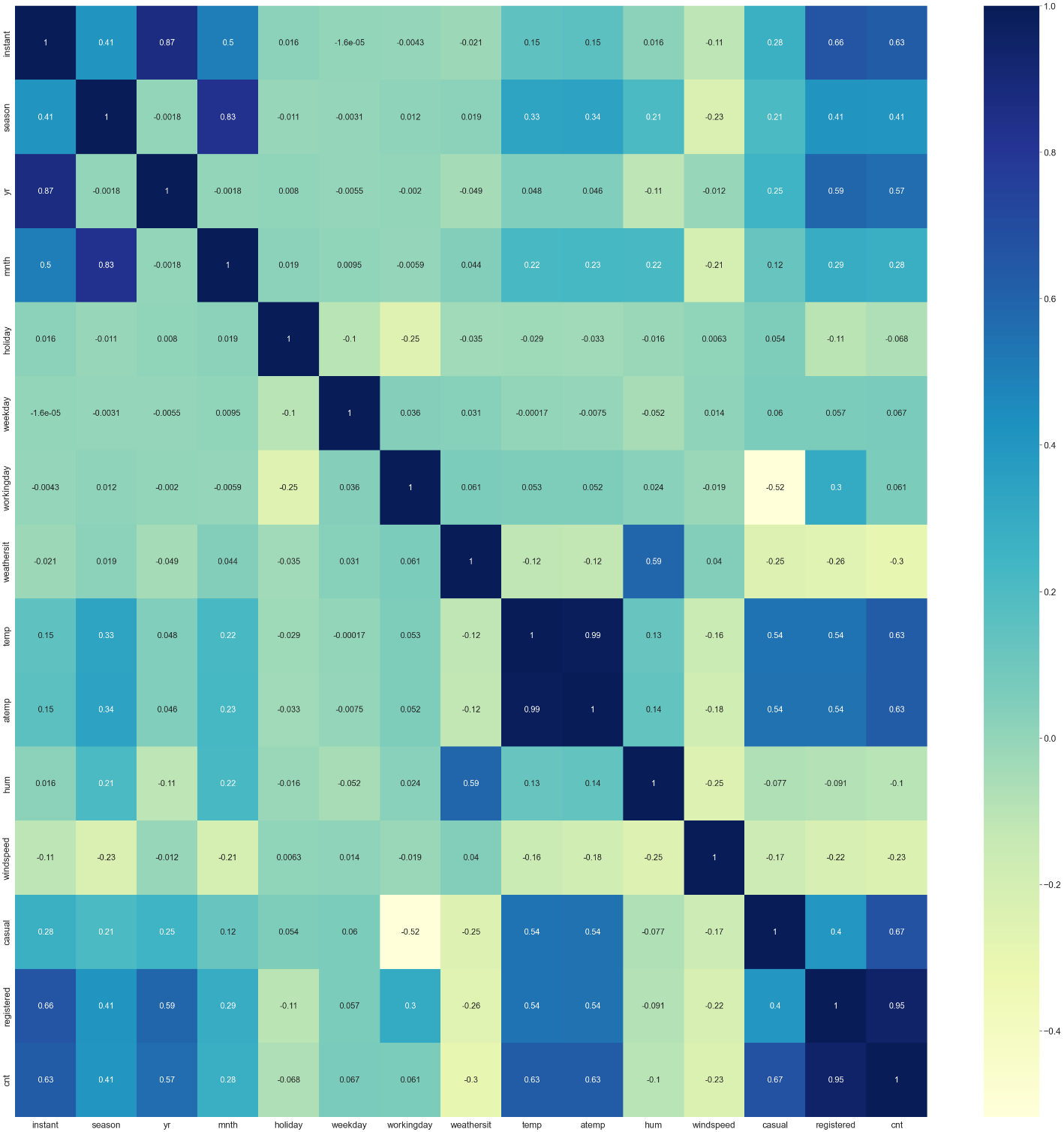
#metode selection (2) : heat-map

corr = day.corr()

sns.set(*rc* = {'figure.figsize':(50,50)}, *font\_scale*=2);

sns.heatmap(corr, *cmap*="YlGnBu", *annot* = True, *linewidths*=0, *annot\_kws*={"size": 20});

plt.show()



Berikut adalah perbandingan hasil dari variable variable yang diperoleh dari Chi-Square dan Heat-Map.

|  |  |  |
| --- | --- | --- |
| No | Chi-Square | Heat-Map |
| 1 | Year | |
| 2 | Month | |
| 3 | Holiday | Season |

|  |  |  |
| --- | --- | --- |
| 4 | Weekday | Temp |

Dari hasil kedua metode tersebut dapat disimpulkan bahwa variable month, year, holiday, season, weekday, dan temp merupakan variable yang memiliki pengaruh yang signifikan.

1. Perbandingan
2. Year

Chart, bar chart

Description automatically generated

Tahun 1 (2012) memiliki total penyewaan sepeda yang jauh lebih banyak dibanginkan dengan Tahun 0 (2011).

1. Month

Chart, bar chart

Description automatically generated

Bulan Mei hingga Oktober memiliki total penyewaan yang lebih banyak dibandingkan dengan bulan bulan lain.

1. Season

Chart, bar chart

Description automatically generated

Musim fall memiliki total penyewaan terbanyak, diikuti oleh musim summer dan musim winter.

1. Holiday

Chart, bar chart

Description automatically generated

Hampir semua dari total penyewaan sepeda, menyewa diluar hari libur atau holiday.

1. Weekday

Chart, bar chart

Description automatically generated

Total penyewaan meningkat dari hari senin hingga jumat, lalu menurun pada hari sabtu hingga minggu, di mana total penyewaan terbanyak terdapat pada hari jumat.

1. Temp

Chart, scatter chart

Description automatically generated

Banyaknya total penyewaan naik pada saat di mana suhu tidak terlalu panas dan tidak terlalu dingin.

[*https://github.com/shrk-sh-ioai/pds-banking-dt/tree/main/tugas-absence/assets*](https://github.com/shrk-sh-ioai/pds-banking-dt/tree/main/tugas-absence/assets)