

Analisis Perbedaan Harga Lyft dan Uber Terhadap Keadaan Cuaca

Jennifer Patricia , Hasri Wiji Aqsari, dan Santi Wulan Purnami
Departemen Statistika, Fakultas Sains dan Analitika Data,
Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail : santiwulan08@gmail.com

Abstrak—Transportasi *online* sedang marak pada beberapa tahun ini. Banyak perusahaan-perusahaan baru yang merintis usaha transportasi *online*. Tetapi ada juga beberapa perusahaan yang memang sudah berdiri sejak lama. Salah satunya Uber dan Lyft. Harga dari transportasi *online* tidak dinamis, ada beberapa faktor yang diduga mempengaruhi dari harga tersebut, diantaranya keadaan cuaca, jarak destinasi, jumlah permintaan dan jumlah ketersediaan mitra pengemudi. Dalam penelitian ini akan divisualisasikan data sehingga didapatkan informasi yang dapat digunakan untuk mengambil kesimpulan. Sebelumnya dilakukan dahulu proses *preprocessing* data untuk memastikan data yang akan divisualisasikan sudah bebas dari *missing value*, *noise*, dan data lain yang mengganggu. Hasil dari visualisasi data didapatkan hasil tidak terdapat perbedaan harga Uber dan Lyft pada kondisi cuaca tertentu. Harga Uber cenderung lebih mahal daripada Lyft di kondisi normal maupun pada kondisi cuaca tertentu.

Kata Kunci: Lyft, *Missing value*, *Preprocessing*, Uber, Visualisasi,

I. PENDAHULUAN

Jasa transportasi memiliki peranan yang sangat penting dalam kehidupan masyarakat, untuk membantu mobilitas dalam mendukung pekerjaan sehari-hari. Transportasi *online* adalah suatu penyelenggaraan lalu lintas dan angkutan jalan yang memanfaatkan perkembangan ilmu pengetahuan (teknologi) berbasis aplikasi dan baik untuk pemesanan maupun pembayaran. Bisnis transportasi *online* sangat menguntungkan, salah satu indikator yang mudah terlihat adalah begitu cepatnya pertumbuhan para pelaku bisnis transportasi atau perusahaan sejenis seperti Go-jek dan Grab jika di Indonesia, Uber dan Lyft di luar negeri.

Kualitas pelayanan dan harga menjadi salah satu faktor penting untuk menjaga dan mempertahankan konsumen agar tidak beralih ke layanan transportasi lainnya. Menurut Lupiyoadi (2013:136), harga memainkan peranan penting dalam bauran pemasaran, karena penentuan harga terkait langsung dengan pendapatan yang diterima oleh perusahaan.

Harga dari transportasi *online* ini sangat dinamis, tergantung dengan keadaan pada hari tersebut. Faktor-faktor yang mempengaruhi harga diantaranya, jarak antara tempat pemesanan dan tempat tujuan, cuaca pada hari tersebut misal hujan atau tidak, jumlah permintaan pelanggan, jumlah ketersediaan mitra pengemudi, dan jenis produk yang dipilih.

Berdasarkan keadaan harga tersebut, banyak perusahaan yang bersaing menawarkan harga terbaik mereka untuk menarik pelanggan. Penulis tertarik melakukan penelitian terhadap harga transportasi *online* bernama Uber dan Lyft. Penulis akan membandingkan harga keduanya

pada kondisi yang normal dan pada kondisi dipengaruhi oleh cuaca di Kota Boston, Amerika.

II. TINJAUAN PUSTAKA

A. *Preprocessing Data*

Preprocessing data adalah salah satu teknik dari data mining yang termasuk persiapan dan transformasi data ke bentuk yang cocok untuk dilakukan prosedur mining. Data *preprocessing* bertujuan untuk mengurangi dimensi data, mencari hubungan antar data, menormalisasi data, mengeluarkan data pencilaan, dan mengekstrak fitur untuk data. Beberapa teknik yang termasuk dalam proses data *preprocessing* adalah pembersihan data, integrasi, transformasi, dan reduksi. Berikut beberapa istilah dalam *preprocessing* data :

1. *Missing values*. Bila ada data yang nilainya tidak tercatat, maka nilai yang hilang ini dapat diisi dengan menggunakan nilai rata-rata.
2. *Binning*. Metode ini digunakan untuk menyimpan data berdasarkan "tetangga", yaitu nilai di sekitarnya. Nilai yang telah diurutkan dibagi ke beberapa kelompok. Metode ini bergantung pada data disekitarnya, sehingga dilakukan *local smoothing*.
3. Reduksi dimensi. Metode ini menggunakan mekanisme pengkodean untuk mengurangi dataset. Reduksi dimensi dapat dilakukan dengan tanpa ada informasi yang hilang. Ada dua tipe reduksi dimensi, yaitu *feature selection* dan *feature extraction*.
4. *Feature selection*. Digunakan untuk menyaring fitur atau variabel yang tidak berhubungan atau mubazir dari dataset. Beberapa metode *feature selection* yaitu :
 - a. *Variance Thresholds*. *Variance thresholds* menghilangkan fitur yang nilainya tidak berubah banyak dari observasi ke observasi.
 - b. *Correlation Thresholds*. *Correlation Thresholds* menghilangkan fitur yang berkorelasi atau berhubungan tinggi dengan fitur lain.

B. *Statistika Deskriptif*

Statistika deskriptif merupakan metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna. Statistika deskriptif dalam praktikum ini terdiri dari *mean*, simpangan baku, median, nilai maksimum dan nilai minimum.

Mean atau rata-rata adalah jumlah semua data yang ada dibagi dengan banyaknya data. Data tunggal memiliki rumus :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (1)$$

Keterangan :

\bar{x} : rata-rata

x_n : sukuk e-n

n : banyaknya data

Simpangan baku menunjukkan penarikan akar dari rata-rata kuadrat jarak suatu data terhadap rata-ratanya. Simpangan baku memiliki rumus :

$$S = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}} \quad (2)$$

Keterangan :

S : simpangan baku

n : banyaknya data

x_i : data ke-i

\bar{x} : rata-rata

Median adalah nilai tengah dari data yang telah disusun berurutan mulai dari data yang terkecil sampai dengan yang terbesar. Secara matematis median dilambangkan dengan *Me* yang dapat dihitung dengan rumus :

$$Me = x_{\frac{n+1}{2}} \quad (3)$$

Keterangan :

Me : median

n : banyaknya data

x : nilai data

Statistik lima serangkai. Nilai minimum, maksimum, dan kuartil memberikan informasi mengenai nilai tengah dan variasi dari sebuah variabel. Ditulis dengan urutan dari yang terkecil hingga terbesar, maka hal ini dikatakan *five-number summary* dari sebuah variabel. Statistik lima serangkai memberikan informasi tentang lokasi dan sebaran suatu data.

$$Min, Q1, Q2, Q3, Max \quad (4)$$

Keterangan :

Min : nilai minimum

Q1 : kuartil bawah

Q2 : median

Q3 : kuartil atas

Max : nilai maksimum

C. Visualisasi Data

Bidang visualisasi difokuskan pada penciptaan gambar yang menyampaikan informasi penting tentang data yang mendasari [1]. Visualisasi data dilihat oleh banyak bidang ilmu sebagai komunikasi visual modern. Visualisasi data tidak berada di bawah bidang manapun, melainkan interpretasi di antara banyak bidang misalnya, terkadang dilihat sebagai cabang modern dari statistik deskriptif oleh beberapa orang, tetapi juga sebagai dasar alat pengembangan oleh yang lain. Visualisasi data mengikutkan pembuatan dan kajian dari representasi visual dari data, artinya informasi yang telah diabstraksikan dalam bentuk skematis, termasuk atribut atau variabel dari unit informasi.

1. Histogram

Data yang telah disusun dalam bentuk tabel distribusi frekuensi dapat disajikan dalam bentuk diagram yang disebut histogram, berikut adalah beberapa pengertian histogram. Histogram yaitu merupakan grafik dari distribusi frekuensi suatu variabel. Tampilan histogram berupa petak-petak empat persegi panjang. Sebagai sumbu horizontal (absis, sumbu x) boleh memakai tepi-tepi kelas, batas-batas kelas atau nilai-nilai variabel yang diobservasi, sedang sumbu vertical (ordinat, sumbu y) menunjukkan frekuensi. Untuk distribusi bergolong/ kelompok yang menjadi absis adalah nilai tengah dari masing-masing kelas [2].

Histogram adalah grafik yang menggambarkan suatu

distribusi frekuensi dengan bentuk beberapa segi empat [3]. Histogram merupakan grafik batang dari distribusi frekuensi [4]. Histogram adalah suatu bentuk grafik yang menggambarkan sebaran (distribusi) frekuensi suatu perangkat data dalam bentuk batang [5]. Histogram digunakan untuk menggambarkan secara visual frekuensi data yang bersifat kontinu.

Jadi histogram adalah diagram kotak yang lebarnya menunjukkan interval kelas, sedangkan batas-batas tepi kotak merupakan tepi bawah dan tepi atas kelas, dan tingginya menunjukkan frekuensi pada kelas tersebut.

2. Scatter Diagram

Scatter diagram adalah gambaran yang menunjukkan kemungkinan hubungan antara keeratan hubungan antara dua variabel tersebut yang sering diwujudkan sebagai koefisien korelasi.

3. Density Plot

Density plot memvisualisasikan distribusi data melalui interval atau periode waktu yang berkesinambungan. Plot ini adalah variasi dari histogram yang menggunakan kernel smoothing untuk memplot nilai, memungkinkan distribusi yang lebih halus dengan menghaluskan kebisingan. Puncak dari density plot membantu menampilkan di mana nilai terkonsentrasi selama interval [6].

4. Correlogram

Correlogram adalah sebuah grafik yang menunjukkan korelasi antara dua variabel.

5. Line Plot

Line plot atau diagram garis dinilai merupakan bentuk yang sangat tepat untuk menggambarkan data yang berhubungan dengan runtutan waktu (*time series data*). Dengan kata lain, corak diagram ini paling sesuai digunakan untuk memvisualisasikan perkembangan dinamika keadaan dari satu kurun waktu ke rentang waktu berikutnya [8]

III. METODE PENELITIAN

A. SumberData

Data yang digunakan dalam penelitian ini merupakan data sekunder yang tersedia di website <https://www.kaggle.com/datasets> dengan menggunakan data *Uber and Lyft* yaitu data tentang perusahaan transportasi *online* Uber dan Lyft di Kota Boston. Diakses pada hari Senin, 2 Maret 2020 pukul 15.01 WIB.

B. Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini adalah sebagai berikut.

Tabel 1 Nama Variabel, Deskripsi dan Tipe Data

| Variabel | Deskripsi | Tipe Data |
|------------------|---|-----------|
| Distance | Jarak dari penjemputan ke tujuan | Numerik |
| Cab_type | Jenis transportasi <i>online</i> (Uber atau Lyft) | Kategorik |
| Time_stamp | Waktu pemesanan | Kategorik |
| Destination | Tempat tujuan | Kategorik |
| Source | Tempat penjemputan | Numerik |
| Price | Harga (USD) | Numerik |
| Surge_multiplier | Kelipatan harga pada waktu tertentu | Numerik |
| Name | Produk yang dihasilkan | Kategorik |
| Temp | Temperatur pada saat kejadian (F) | Numerik |
| Clouds | Bagian langit yang tertutup awan | Numerik |
| Pressure | Tekanan pada saat kejadian (mb) | Numerik |
| Rain | Satuan hujan pada 1 jam terakhir (inch) | Numerik |
| Humidity | Kelembapan pada saat | Numerik |

| Wind | kejadian (%) Kecepatan angin (mph) | Numerik |
|------|---------------------------------------|---------|
|------|---------------------------------------|---------|

C. Langkah Analisis

Langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut.

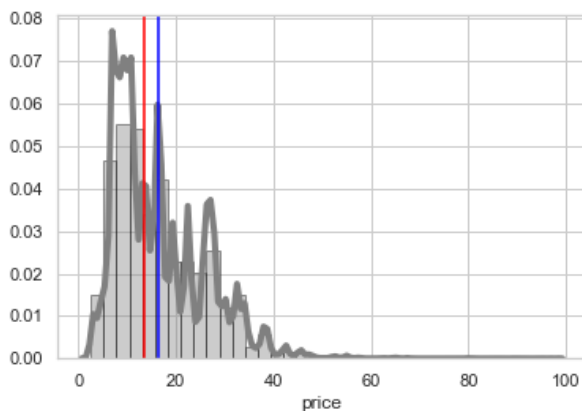
1. Mengunduh data dari website <https://www.kaggle.com/datasets>
2. Melakukan *preprocessing* data
3. Melakukan visualisasi data
4. Memberikan deskripsi hasil *preprocessing* dan visualisasi data
5. Memberikan saran metode statistika yang sesuai dengan data.

IV. HASIL DAN PEMBAHASAN

A. Hasil Preprocessing Data

Preprocessing yang telah dilakukan menghasilkan beberapa keadaan, diantaranya :

1. Menggabungkan data *cab_rides* dengan data *weather*. Data tersebut digabungkan berdasarkan variabel *source* yaitu tempat penjemputan, tanggal dan waktu penjemputan. Penggabungan data dilakukan karena akan dilihat apakah ada pengaruh antara harga dengan keadaan cuaca, tetapi harga dan cuaca berada di file yang berbeda.
2. Memeriksa data apakah ada yang *missing value* atau tidak. Hasilnya pada variabel *price* terdapat sebanyak 100494 data *missing value*. Pada variabel *rain* terdapat sebanyak 1028034 data *missing value*.
3. Membuat *density plot* dengan garis *mean* dan median untuk penentuan pengisian *missing value*. *Density plot* ditampilkan pada gambar berikut.



Gambar 1 Density plot

Berdasarkan gambar tersebut didapatkan hasil untuk variabel *price* diisi dengan *mean*, karena *mean* lebih berada di tengah distribusi daripada median. Sedangkan variabel *rain* diisi dengan nilai nol, dengan artian tidak ada hujan.

4. Melakukan *binning data* yaitu mengkategorikan data numerik. Data yang dikategorikan yaitu pada variabel *rain*. Dengan ketentuan nilai -1 sampai 0.000001 dikategorikan sebagai tidak hujan, dan 0.000001 sampai 0.8 dikategorikan sebagai hujan.
5. Melakukan *feature selection* dengan metode *variance threshold*. Hasil dari metode tersebut yaitu mengeluarkan variabel *humidity* dan *clouds* karena memiliki variansi yang kecil, sehingga diduga data memiliki nilai yang hampir sama. Metode lain yang

dilakukan yaitu *correlation threshold*, hasil yang didapatkan variabel *time_stamp_x* memiliki korelasi yang tinggi dengan variabel *time_stamp_y* sehingga salah satunya bisa dikeluarkan.

6. Melakukan *convert variable* pada variabel *time_stamp_x*. Nilai pada variabel *time_stamp_x* diubah bentuk menjadi nilai waktu.

B. Statistika Deskriptif

Statistika deskriptif yang dihasilkan dari data tersebut ditampilkan pada tabel berikut :

Tabel 2 Statistika Deskriptif

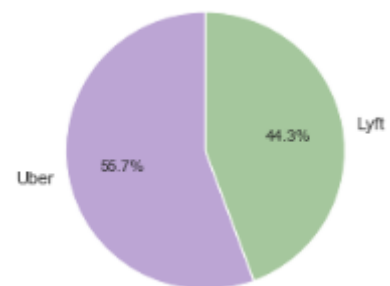
| Varia ble | Mea n | Max | Min | Q1 | Q2 | Q3 |
|---------------------|----------|---------|-------|--------|---------|---------|
| <i>Distance</i> | 2.19 | 7.86 | 0.02 | 1.27 | 2.15 | 2.93 |
| <i>Price</i> | 16.53 | 97.5 | 2.5 | 9.5 | 16 | 22.5 |
| <i>Surge</i> | | | | | | |
| <i>– multiplier</i> | 1.01 | 3 | 1 | 1 | 1 | 1 |
| <i>Temp</i> | 39.5 | 55.06 | 19.78 | 37.34 | 39.59 | 43.69 |
| <i>Pressure</i> | 1005.04 | 1035.07 | 98.52 | 993.96 | 1001.72 | 1014.57 |
| <i>Wind</i> | 7.63 | 18.18 | 0.3 | 5.15 | 8.44 | 10.23 |

Informasi yang didapatkan dari tabel 1 yaitu, data diambil pada musim dingin di Kota Boston, karena dilihat dari nilai temperature yang sangat rendah. Variabel harga memiliki variasi yang sangat tinggi, dapat dilihat dari maksimal dan minimal data yang memiliki range besar. Dapat dilihat bahwa pada data, rata-rata jauh perjalanan adalah 2.19 miles. Sehingga dapat disimpulkan bahwa rata-rata jarak berpergian pengguna transportasi online baik Uber maupun Lyft tidak terlalu jauh.

C. Visualisasi Data

Data setelah dilakukan *preprocessing* selanjutnya divisualisasikan sehingga dapat memberikan informasi. Visualisasi data yang dihasilkan diantaranya :

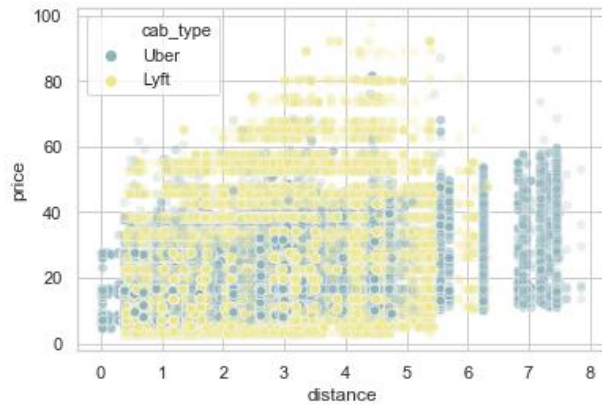
1. *Pie chart* untuk pengguna Uber dan Lyft



Gambar 2 Pie chart pengguna Uber dan Lyft

Informasi yang didapatkan dari gambar di atas adalah berdasarkan data yang diolah, lebih banyak yang pengguna transportasi online Uber daripada Lyft yaitu dengan proporsi sebesar 56.7% dan 44.3% secara berurutan.

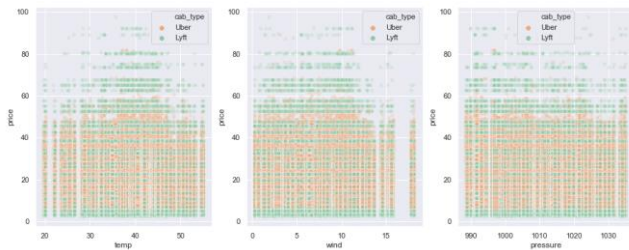
2. *Scatterplot* untuk hubungan variabel *price* dan variabel *distance* pada Uber dan Lyft



Gambar 3 *Scatterplot* variabel *price* dan *distance*

Informasi yang didapatkan dari gambar di atas yaitu transportasi *online* Uber memberikan harga yang murah kepada pengguna daripada Lyft. Selain itu harga di Uber cenderung lebih stabil dengan jarak yang dekat maupun jauh. Sedangkan di Lyft harga cenderung lebih bervariasi untuk jarak dekat maupun jauh. Sehingga dapat disimpulkan dari diagram *scatterplot* di atas bahwa tidak ada hubungan yang kuat antara harga dengan jarak untuk transportasi *online* Uber maupun Lyft. Namun, pengguna transportasi *online* lebih memilih menggunakan Uber ketika perjalanan jarak jauh, yaitu lebih dari 6 miles.

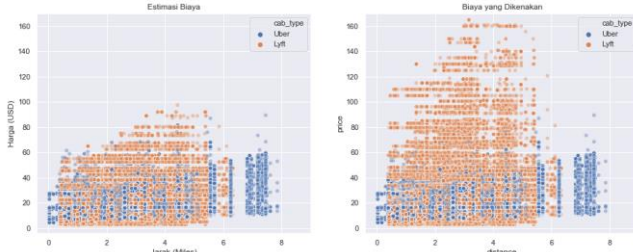
3. *Scatterplot* antara Variabel Harga dengan Variabel Cuaca



Gambar 4 *Scatterplot* untuk harga dan cuaca

Berdasarkan grafik *scatterplot* di atas dapat disimpulkan bahwa tidak ada hubungan antara harga dengan keadaan cuaca. Keadaan cuaca yang dimaksud adalah temperatur, kecepatan angin dan tekanan udara. Dari diagram di atas dapat dilihat bahwa tidak ada pola yang jelas untuk hubungan antara kedua variabel tersebut. Untuk setiap nilai tertentu pada variabel temperature, kecepatan angin, dan tekanan udara memiliki nilai harga yang berbeda-beda. Selain itu berdasarkan gambar tersebut dapat dijelaskan bahwa harga Lyft cenderung lebih mahal daripada Uber di setiap keadaan cuaca yang bermacam-macam.

4. *Scatterplot* antara Variabel Harga dengan Jarak



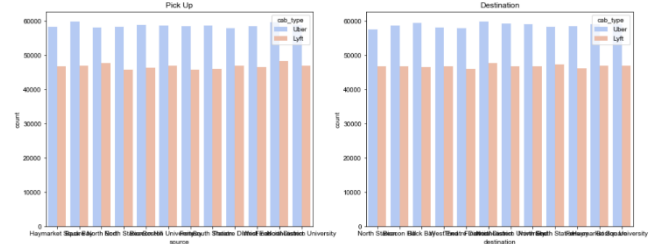
Gambar 5 *Scatterplot* untuk harga dan jarak

Diagram *scatterplot* menampilkan hubungan antara jarak dan harga pada keadaan harga normal dan pada saat harga mengalami kenaikan. Adanya kenaikan harga ini dapat dilihat dari variabel *surge multiplier*, dimana semakin besar nilainya artinya harga mengalami kenaikan yang semakin tinggi. Ada beberapa faktor yang mempengaruhi terjadinya kenaikan harga ini yang nanti akan dibahas lebih

lanjut.

Informasi yang didapatkan dari gambar diatas yaitu harga Uber cenderung lebih stabil dan memiliki nilai harga yang hampir sama pada harga normal dan harga yang sudah mengalami kenaikan. Artinya pada data yang diolah, harga dari transportasi *online* Uber tidak pernah mengalami kenaikan dalam kondisi apapun. Sedangkan harga untuk transportasi *online* Lyft lebih sering mengalami kenaikan. Terbukti dengan bila dibandingkan antara kedua grafik *scatterplot* di atas, data untuk Lyft pada grafik biaya yang dikenakan memiliki nilai yang jauh lebih tinggi daripada pada grafik estimasi biaya.

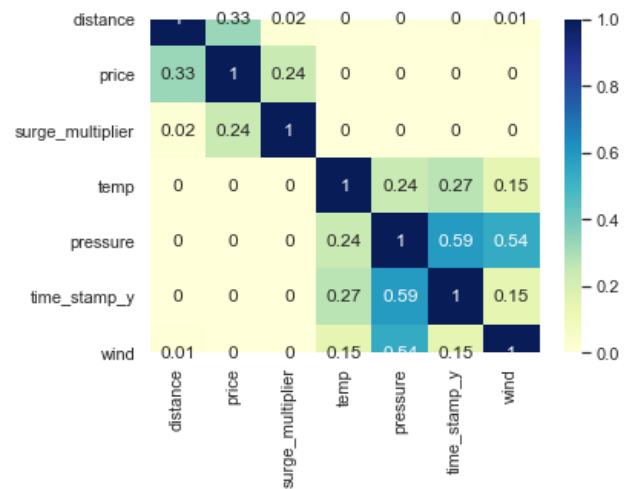
5. *Bar Chart* jumlah penjemputan dan tujuan



Gambar 6 *Bar chart* jumlah penjemputan dan tujuan

Informasi yang didapatkan dari gambar diatas yaitu jumlah pemesanan transportasi *online* Uber maupun Lyft di setiap titik penjemputan hampir sama. Namun dapat disimpulkan bahwa jumlah pemesanan Uber lebih banyak daripada jumlah pemesanan Lyft di setiap titik penjemputan. Pada titik tujuan juga sama, jumlah pemesanan Uber hampir sama pada setiap titik tujuan. Begitu juga dengan jumlah pemesanan Lyft. Sehingga dapat disimpulkan bahwa pemesanan transportasi *online* berdasarkan data untuk titik penjemputan dan pengantaran terbagi secara rata.

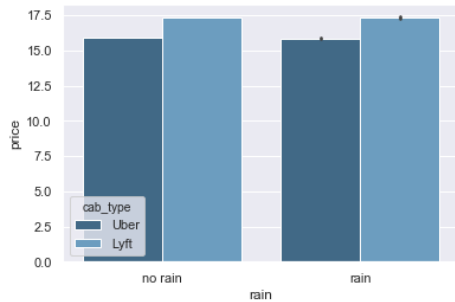
6. Korelogram untuk Setiap Variabel Numerik



Gambar 7 Korelogram untuk setiap variabel

Berdasarkan grafik korelogram di atas, dapat dilihat bahwa tidak ada hubungan sama sekali antara temperatur, tekanan udara, dan kecepatan angin terhadap harga transportasi *online* yang terbukti dari nilai korelasi sebesar 0. Sehingga dapat disimpulkan bahwa keadaan cuaca tidak mempengaruhi naik turunnya harga transportasi *online* untuk Uber maupun Lyft. Namun ada hubungan positif yang renggang antara harga dengan jarak dibuktikan dari nilai korelasinya sebesar 0,33. Juga ada hubungan positif yang renggang antara variabel *surge multiplier* dengan harga.

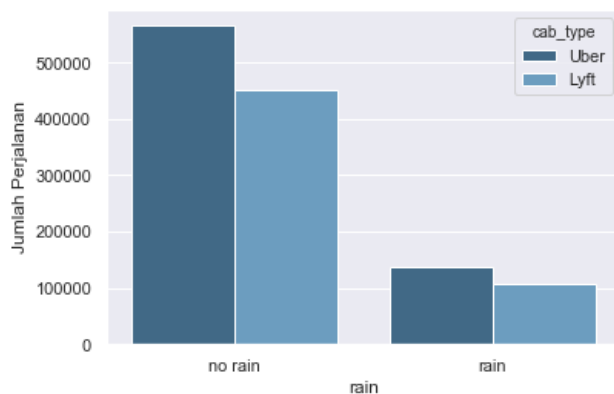
7. Bar chart antara Variabel Harga dengan Variabel Hujan



Gambar 8 Bar chart variabel hujan dan harga

Dapat dilihat dari grafk di atas, keadaan hujan tidak mempengaruhi harga transportasi *online* untuk Uber maupun Lyft. Rata-rata harga transportasi *online* Uber pada saat hujan dan tidak hujan hampir sama yaitu sekitar 15 USD. Sedangkan untuk transportasi *online* Lyft pada saat hujan dan tidak hujan juga hampir sama yaitu di sekitar 17 USD.

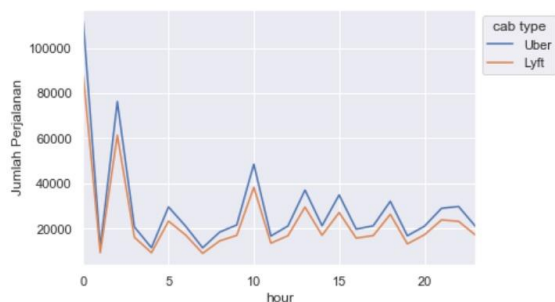
8. Bar chart untuk variabel hujan dan jumlah perjalanan



Gambar 9 Bar chart hujan dan jumlah perjalanan

Informasi yang didapatkan dari gambar diatas yaitu berdasar data, keadaan hujan memiliki jumlah perjalanan yang lebih besar pada transportasi *online* Uber maupun Lyft. Pada keadaan tidak hujan jumlah pemesanan jauh lebih banyak daripada jumlah pemesanan ketika hujan. Hal tersebut terjadi di Uber maupun Lyft. Pada keadaan tidak hujan jumlah pemesanan Uber berada pada jumlah 500.000 sedangkan saat hujan jumlah pemesanan ada pada jumlah 100.000. Terjadi perbedaan yang sangat jauh.

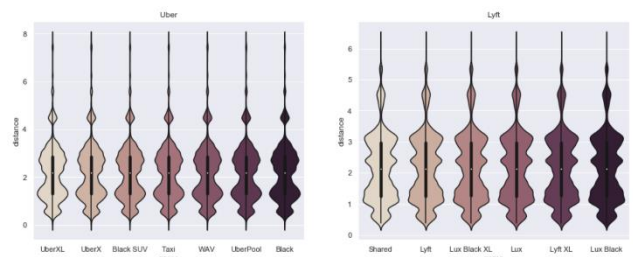
9. Line chart variabel waktu dan jumlah perjalanan



Gambar 10 Line chart variabel waktu dan jumlah perjalanan

Dari gambar *line chart* diatas dapat diketahui bahwa tiap jam memiliki jumlah perjalanan transportasi *online* antara Uber dan Lyft yang selalu berhimpitan. Puncak jumlah perjalanan tertinggi ada pada jam 12 malam, yaitu mencapai lebih dari 80.000 untuk Lyft dan lebih dari 100.000 untuk Uber. Jumlah perjalanan paling rendah berada ketika jam 1 pagi, yaitu sebanyak kurang dari 10.000 perjalanan antara Uber maupun Lyft.

10. Violin plot variabel jarak untuk tiap produk



Gambar 11 Violin plot variabel jarak untuk tiap produk

Dari gambar *violin plot* diatas dapat diketahui bahwa tidak ada perbedaan antara jarak yang ditempuh dengan pemilihan produk untuk transportasi *online* Uber maupun Lyft. Range jarak pada Uber lebih besar daripada range jarak pada Lyft. Jarak maksimal pada Uber sekitar 8 miles sedangkan jarak maksimal pada Lyft sekitar 6 miles.

11. Tabel perbandingan transportasi *online*

Tabel 3 Statistika Deskriptif

| | cab_type | price | distance | temp |
|---|----------|-----------|----------|-----------|
| 0 | Lyft | 17.333392 | 2.183744 | 39.498194 |
| 1 | Uber | 15.891444 | 2.191278 | 39.498135 |

Berdasarkan tabel tersebut dapat diketahui bahwa transportasi *online* Uber dan Lift memiliki rata-rata jarak perjalanan yang hampir sama yaitu bernilai 2.1837 *miles* dan 2.1912 *miles* secara berurutan. Rata-rata harga antara transportasi *online* untuk kedua perusahaan tersebut memiliki sedikit perbedaan, yaitu 17.33 USD untuk Lyft dan 15.89 USD untuk Uber tanpa melihat kondisi cuaca. Perjalanan yang ditempuh oleh kedua transportasi *online* berada di kondisi cuaca yang mirip dengan rata-rata temperature sebesar 39.498 *Fahrenheit* untuk keduanya.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

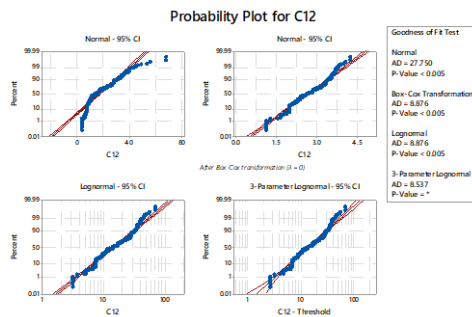
Berdasarkan hasil visualisasi data pada bab sebelumnya didapatkan kesimpulan sebagai berikut :

1. Pengguna transportasi *online* Uber lebih banyak daripada Lyft yaitu dengan proporsi sebesar 56.7% dan 44.3% secara berurutan
2. Transportasi *online* Uber memberikan harga yang murah kepada pengguna daripada Lyft. Selain itu harga di Uber cenderung lebih stabil dengan jarak yang dekat maupun jauh. Sedangkan di Lyft harga cenderung lebih bervariasi untuk jarak dekat maupun jauh.
3. Tidak ada hubungan antara harga dengan keadaan cuaca. Keadaan cuaca yang dimaksud adalah temperatur, kecepatan angin dan tekanan udara. Harga Lyft cenderung lebih mahal daripada Uber di setiap keadaan cuaca yang bermacam-macam.
4. Harga Uber cenderung lebih stabil dan memiliki nilai harga yang hampir sama pada harga normal dan harga yang sudah mengalami kenaikan. Sedangkan harga untuk transportasi *online* Lyft lebih sering mengalami kenaikan
5. Keadaan hujan tidak mempengaruhi harga transportasi *online* untuk Uber maupun Lyft. Keadaan hujan memiliki jumlah perjalanan yang lebih besar pada transportasi *online* Uber maupun Lyft.
6. Puncak jumlah perjalanan tertinggi ada pada jam 12 malam, sedangkan perjalanan terendah ada pada jam 1 pagi

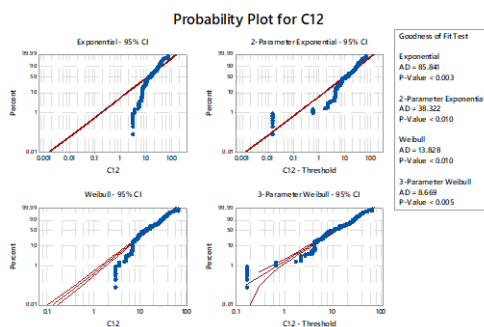
B. Saran

Saran untuk penelitian lanjutan yaitu melakukan uji hipotesis *mean* dua populasi independen, dimana populasi pertama adalah harga untuk transportasi *online* Uber dan populasi kedua adalah harga untuk transportasi *online* Lyft. Uji hipotesis *mean* dua populasi independen bertujuan untuk mengetahui apakah rata-rata harga dari transportasi *online* Uber lebih murah atau lebih mahal daripada transportasi *online* Lyft.

Berikut dilakukan pengujian untuk mengetahui distribusi dari variabel *price*. Hal ini dilakukan untuk mengetahui jenis regresi yang cocok untuk diterapkan pada data ini.



Gambar 12 Output pengujian data berdistribusi lognormal



Gambar 13 Output pengujian data berdistribusi wibull

Berdasarkan gambar tersebut, didapatkan *p-value* untuk uji distribusi lognormal yaitu < 0.005 , sedangkan *p-value* untuk uji distribusi Weibull yaitu < 0.010 . Sehingga dapat disimpulkan bahwa variabel *price* mengikuti distribusi lognormal. Saran yang penulis berikan untuk penelitian selanjutnya, jenis regresi yang cocok digunakan adalah regresi lognormal. Variabel respon yang digunakan dalam model regresi adalah variabel harga. Karena peneliti ingin mengetahui hubungan antara harga dengan variabel lain. Variabel prediktor yang dapat digunakan dalam model regresi adalah jenis transportasi *online*, jarak dan *surge_multiplier*. Karena berdasarkan dari grafik korelogram sebelumnya, ketiga variabel tersebut adalah variabel yang memiliki hubungan dengan variabel harga.

DAFTAR PUSTAKA

- [1] Hansen, C. D. & Johnson, C. R. 2005. The Visualization handbook. USA: Elsevier Inc.
- [2] Somantri, Ating dan Sambas Ali Muhidin. 2006. Aplikasi statistika dalam Penelitian. Bandung : Pustaka Ceria
- [3] Riduwan . 2010. Dasar-dasar Statistika. Bandung : Alfabeta.
- [4] Hasan, M. Iqbal. 2011. Pokok – Pokok Materi Statistika (Statistik Deskriptif). Jakarta : PT Bumi Aksara.
- [5] Furqon. 1999. Statistika Terapan Untuk Penelitian.

AFABETA: Bandung.

- [6] Website : <http://www.visualcomplexity.com/> diakses pada tanggal 02 Maret 2020
- [7] Website : : <http://datavisualization.ch/showcases/> diakses pada tanggal 02 Maret 2020
- [8] Santosa, P. B., & Hamdani, M. (2007). *Statistika* Edisi Kedua. Jakarta: Remaja Rosdakarya

