

BAB I

PENDAHULUAN

1.1.Latar Belakang

Perkembangan teknologi informasi dan komunikasi pada abad ke 21 menjadi tanda adanya peningkatan penggunaan teknologi yang semakin maju. Kebutuhan informasi pada masa ini tidak hanya pada masyarakat kelas-kelas tertentu, akan tetapi semua elemen masyarakat dapat memanfaatkan informasi yang tanpa batas. Masyarakat dapat mengakses setiap informasi dalam bentuk jejaring internet secara online. Dengan keterbukaan dan keterluasan informasi yang terdapat di internet maka muncul budaya baru seperti penggunaan media sosial/virtual sebagai bentuk pemanfaatan informasi dan interaksi dalam berkomunikasi. Dalam artian untuk mendapatkan suatu informasi tentulah para pengguna akan melihat terhadap akun-akun yang ada di media sosial seperti facebook, twitter, instagram, tiktok dan sebagainya, kemudian dari beberapa akun media sosial tersebut seorang pengguna akan mengikuti tautan yang terdapat dalam akun pilihannya. Dari alur ini maka muncul suatu metrics untuk mengukur kreadibilitas akun media sosial tersebut terhadap informasi yang dimanfaatkan oleh para pengguna media sosial. (Bronman,2014).

Robin & Rachel (2015) mendefenisikan *altmetrics* (data metrik) merupakan sebuah studi baru tentang pengukuran berdasarkan pemanfaatan media sosial dalam melakukan proses analisa dan menginformasikan suatu artikel ilmiah. Perkembangan altmetrics yang dimulai dengan munculnya world wide web pada tahun 1990 dan munculnya facebook pada tahun 2004 dan pada tahun 2008 disusul dengan hadirnya penyempurnaan pencarian informasi ilmiah melalui *academic network*, hingga terciptanya *open access* sebuah jurnal ilmiah melalui tautan akademik tersebut. *Altmetrics* dirasa perlu untuk mengukur kebutuhan pembaca melalui sosial media dan hadir pada tahun 2010. Dalam penelitian ini data yang akan digunakan yaitu data metrik kinerja *Facebook* yaitu data yang terkait dengan postingan yang dipublikasikan sepanjang tahun 2014 di laman

Facebook salah satu merek kosmetik ternama. *Dataset* ini berisi 500 dari 790 baris dengan jumlah variabel 19 (Sembilan belas) variabel yaitu 18 (Delapan belas) variabel bebas (X) dan 1 (satu) variabel terikat (Y) (Moro et al. 2016).

Regresi linear (*linear regression*) merupakan metode yang digunakan untuk memperoleh model hubungan antara satu variabel terikat dengan satu variabel bebas. Harlan (2018:5) menyebutkan bahwa Variabel terikat pada regresi linear disebut juga sebagai respons, sedangkan variabel bebas dikenal sebagai prediktor atau regresor. Jika hanya terdapat satu variabel terikat dalam model maka teknik ini disebut sebagai regresi linear sederhana (*simple linear regression*), sedangkan jika terdapat beberapa variabel bebas, teknik ini disebut regresi linear berganda (*multiple linear regression*). Analisis regresi memiliki 3 kegunaan, yaitu untuk tujuan deskripsi fenomena data atau kasus yang sedang diteliti, untuk tujuan kontrol dan untuk tujuan prediksi.

Bawono dan Shina (2018:20) mengatakan model regresi linier disebut valid apabila model tersebut memenuhi asumsi klasik. uji asumsi klasik merupakan uji prasyarat yang dilakukan sebelum melakukan analisis lebih lanjut terhadap data yang telah dikumpulkan. Pengujian asumsi klasik ini ditujukan agar dapat menghasilkan model regresi yang memenuhi kriteria BLUE (*Best Linier Unbiased Estimator*) yang digunakan sebagai estimator terpercaya, dimana estimator tersebut dinyatakan tidak bias, konsisten, berdistribusi normal dan juga efisien. Untuk mengetahui apakah model regresi yang akan digunakan telah memenuhi kriteria BLUE maka perlu dilakukan serangkaian pengujian yaitu Uji normalitas, uji multikolinieritas, uji heteroskedastisitas, dan uji autokorelasi. Ghozali (2016) menyatakan bahwa normalitas terjadi ketika variabel pengganggu atau residual dalam persamaan regresi memiliki distribusi normal, multikolinieritas terjadi ketika ada korelasi diantara variabel bebas, heteroskedastisitas terjadi jika varian dari *error* suatu pengamatan ke pengamatan lain terjadi ketidaksamaan (tidak konstan), sedangkan autokorelasi terjadi ketika nilai korelasi dalam persamaan regresi bernilai nol. Asumsi yang akan dibahas dalam penelitian ini adalah asumsi multikolinieritas.

Kusrini Setiawan (2010) mendefinisikan bahwa multikolinearitas merupakan adanya hubungan korelasi linear yang tinggi diantara beberapa atau semua variabel bebas dari model regresi berganda. Multikolinearitas terjadi apabila terdapat hubungan linear diantara beberapa variabel atau seluruh variabel bebas. Semakin tinggi multikolinearitas antara variabel, semakin tinggi galat koefisien regresinya. Galat yang tinggi akan menghasilkan penduga parameter yang bias. Untuk mengatasi multikolinearitas, salah satu metode yang dapat digunakan adalah PCA (*Principal Component Analysis*) atau sering disebut dengan metode analisis komponen utama.

Analisis Komponen Utama (PCA) adalah metode untuk membangun beberapa atribut baru yang merupakan kombinasi linier dari atribut-atribut asli (Sammut 2011). Metode PCA sangat berguna digunakan jika data yang ada memiliki jumlah variabel yang banyak dan terdapat korelasi antar variabel. Perhitungan PCA didasarkan pada perhitungan nilai eigen dan vektor eigen yang merepresentasikan sebaran data dari suatu dataset. Metode ini ditemukan pertama kali oleh Karl Pearson pada tahun 1901, dan dikembangkan oleh Harold Hotelling pada tahun 1933. Kemudian pada tahun 1947 teori tersebut diperkenalkan kembali oleh Karhunen, dan dikembangkan oleh Loeve pada tahun 1963 (Jolliffe 2002).

Dalam analisis komponen utama akan menghasilkan variabel-variabel baru yang merupakan kombinasi linear dari variabel bebas asal dan antar variabel baru yang bersifat saling bebas dan juga akan menghasilkan estimasi titik. Untuk mengukur keakuratan estimasi titik tersebut, maka dapat digunakan metode bootstrap yang menghasilkan interval kepercayaan dengan cara resampling data dengan pengembalian. *Bootstrap* adalah metode re-sampling yang berlaku umum untuk memperkirakan sifat ketidakpastian dari suatu statistik seperti bias, kesalahan standar atau batas kepercayaan. *Bootstrap* memperkirakan kesalahan standar statistik dari data empiris (asli) dengan pengambilan sampel ulang dan tidak mengasumsikan distribusi apa pun untuk perkiraan ketidakpastian, oleh karena itu dapat diterapkan untuk semua data statistik. (Zhang et al., 2017)

Sahinler dan Topuz (2019) menyatakan bahwa *bootstrap* adalah Teknik *resampling* nonparametrik yang bertujuan untuk menentukan estimasi standar error dan interval kepercayaan (*interval confidensi*) dari parameter populasi tanpa mengetahui sebaran data (tanpa sumsi distribusi). *bootstrap* dapat digunakan untuk mengatasi permasalahan dalam statistika baik masalah data yang sedikit maupun data yang tidak memiliki asumsi dalam distribusinya, karena pengambilan sampel *bootstrap* dilakukan berulang kali maka kemampuan komputer yang tinggi diperlukan ,sehingga *bootstrap* dikatakan juga metode berbasis komputer.

Penelitian sebelumnya, Ryan, Iskandar dkk (2013) mengkaji perbandingan metode *bootstrap* dan metode *jackknife* dalam menaksir parameter regresi untuk mengatasi multikolinearitas, data yang digunakan adalah seratus ekor ikan yang sama spesiesnya dari studi perikanan di Universitas Mustafa Kemal (Turkey) dimana panjang sirip ikan dan panjang ekor ikan tidak saling berkorelasi sebagai variabel bebasnya yang menjelaskan variasi umur ikan. Hasil penelitiannya menunjukkan bahwa bias parameter, standar error, dan interval konfidensi Jackknife lebih besar dibandingkan Bootstrap. Sehingga metode Bootstrap lebih efisien dalam menduga parameter regresi dibandingkan metode Jackknife ketika terjadi multikolinearitas.

Berdasarkan uraian latar belakang di atas, penulis tertarik ingin melakukan penelitian mengenai “Estimasi *Bootstrap* Untuk Interval Kepercayaan Parameter Model Regresi Berganda Dengan Adanya Multikolinearitas Menggunakan *Principal Component Analysis* ”

1.1 Rumusan Masalah

Dari latar belakang masalah diatas yang menjadi rumusan masalah adalah :

1. Bagaimana Penggunaan metode PCA (*Principal Component Analysis*) dalam mengatasi masalah Multikolinearitas yang terjadi pada data metrik *Facebook*?

2. Bagaimana penggunaan metode PCA (*Principal Component Analysis*) dengan Bootstrap dalam mengatasi masalah multikolinearitas yang terjadi pada data metrik facebook ?

1.2 Batasan Masalah

Untuk menjaga agar penelitian ini lebih terarah dan tidak meluas, maka peneliti membuat batasan masalah sebagai berikut.

1. Masalah yang ingin diatasi adalah masalah multikolinearitas dengan menganggap bahwa asumsi klasik lain terpenuhi.
2. Metode yang digunakan dalam mengatasi masalah multikolinearitas adalah metode PCA (*Principal Component Analysis*) dengan *Bootstrap*.
3. Data yang digunakan adalah data Sekunder yaitu data metrik *facebook* yang diperoleh melalui website *UCI Machine learning* .

1.3 Tujuan Penelitian

Adapun Tujuan Dari Penelitian ini adalah sebagai Berikut

1. Mengimplementasikan penggunaan metode PCA (*Principal Component Analysis*) dalam mengatasi masalah multikolinearitas yang terjadi pada data metrik *facebook*.
2. Mengimplementasikan penggunaan metode PCA (*Principal Component Analysis*) dengan *Bootstrap* dalam mengatasi masalah multikolinearitas yang terjadi pada data metrik *facebook* .

1.4 Manfaat Penelitian

Dengan diadakannya penelitian ini diharapkan dapat memberi manfaat sebagai berikut:

1. Bagi penulis, penelitian ini berguna untuk menambah wawasan mengenai penggunaan metode *PCA (Principal Component Analysis)* dalam mengatasi masalah multikolinearitas yang terjadi pada data metrik *facebook*.

2. Bagi para pembaca, sebagai tambahan informasi dan referensi bacaan mengenai penggunaan PCA (*Principal Component Analysis*) dengan Bootstrap bagi yang hendak melakukan penelitian serupa.
3. Bagi akademik, memberikan tambahan ilmu dan wawasan yang baru tentang cara mendeteksi dan mengatasi masalah multikolinearitas menggunakan PCA (*Principal Component Analysis*) dengan *Bootstrap*.

