# BAB IITINJAUAN PUSTAKA

## Penelitian Terdahulu

Penulis mengambil beberapa referensi yang berhubungan dengan penelitian ini, diantaranya adalah penelitian yang berjudul “Fenomena Selebgram Sebagai Bentuk Personal Branding” (Anindita, 2016) yang membahas tentang fenomena selebriti Instagram sebagai bentuk *Personal Branding* sebagai upaya dari para selebgram untuk dikenal orang sebagai sosok unik dan kreatif melalui aktivitas foto yang diunggah di Instagram. Serta keinginan untuk menyalurkan hobi dan keterampilan melalui akun Instagramnya yang memberikan manfaat berupa mendapatkan endorsement, teman baru, relasi, uang dan pekerjaan. Penelitian kedua berjudul “*Machine Learning Techniques for Brand-Influencer Matchmaking on the Instagram Social Network”*(Sweet et al., 2019) menjelaskan bahwa konten unggahan disertai dengan keterangan atau *caption* yang jelas, dapat mempermudah industri untuk memilih suatu akun untuk dijadikan kerjasama, serta diterapkannya metode algoritma K-Nearest Neighbor untuk kebutuhan klasifikasi ketepatan antara presentasi diri dengan konten yang diposting karena terbukti efektif. Penelitian selanjutnya dengan judul “*Characterising and Detecting Sponsored Influencer Posts on Instagram”*(Zarei et al., 2020) yang lebih fokus kepada ciri-ciri dan karakteristik sebuah akun Instagram yang diduga akun influencer dan juga membahas klasifikasi jenis influencer menggunakan Deep Neural Network(DNN) dengan bantuan beberapa parameter seperti caption, jumlah like dan comment, jumlah follower, biography, dan juga keterangan akun tersebut sudah diverifikasi oleh pihak Instagram. Pada penelitian terakhir yang membuktikan dan menyimpulkan bahwa algoritma K-Nearest Neighbor paling banyak digunakan untuk klasifikasi dengan mengelompokkan data, kemiripan, atau kedekatan yang ada pada *data training*. Pada penelitian yang berjudul “Machine Learning Prediksi Karakter Pengguna Hashtag(#) Bahasa Generasi Milenial di Sosial Media”(Sinaga, 2020) ini juga dijelaskan proses pengolahan data mentah hingga menjadi data load yang siap untuk diujikan.

## Teori Terkait

### Media Sosial

Media sosial adalah medium di internet yang memungkinkan pengguna merepresentasikan dirinya maupun berinteraksi, bekerja sama, berbagi, berkomunikasi dengan pengguna lain membentuk ikatan sosial secara virtual. Dalam media sosial, tiga bentuk yang merujuk pada makna bersosial adalah pengenalan (*cognition*), komunikasi (*communicate*) dan kerjasama (*co-operation*)(Nasrullah, 2015). Beberapa contoh media sosial yang ada saat ini adalah Instagram, Facebook, Twitter, Youtube, dan masih banyak lagi.

### Instagram

Instagram merupakan aplikasi yang digunakan untuk membagikan foto dan video. Instagram sendiri masih merupakan bagian dari Facebook yang memungkinkan teman Facebook itu mengikuti kita dalam akun sosial media Instagram. Karena termasuk salah satu media sosial yang populer untuk membagikan foto dan video, banyak pengguna Instagram yang terjun ke ranah bisnis seperti akun sosial bisnis yang dapat digunakan melakukan produk atau layanann yang disediakan lewat Instagram(Nisrina, 2015). Media sosial Instagram memiliki beberapa fitur yang bisa digunakan, diantaranya sebagai berikut :

1. *Follower*

Sistem sosial yang ada di aplikasi Instagram adalah dengan menjadi pengikut dari akun pengguna lainnya, atau memiliki pengikut Instagram. Dengan demikian komunikasi antara sesama pengguna Instagram sendiri dapat terjalin dengan memberikan tanda suka dan juga mengomentari foto-foto yang telah diunggah oleh pengguna lainnya. Pengikut juga menjadi salah satu unsur yang penting, dimana jumlah tanda suka dari pengikut sangat mempengaruhi apakah foto tersebut dapat menjadi sebuah foto yang disukai atau tidak.

1. Unggah Foto

Fitur utama dalam Instagram adalah sebagai tempat untuk mengunggah dan berbagi foto dan video kepada pengguna lainnya. Foto atau video yang ingin diunggah dapat diperoleh dari kamera ataupun *file* yang sudah ada di album handphone pengguna Instagram.

1. Keterangan/*Caption* Foto

Foto yang telah diunggah dapat diberikan penjelasan tentang foto tersebut dan dapat diberikan penjelasan dimana lokasi foto tersebut diambil.

1. Tanda Suka

Instagram memiliki fitur tanda suka yang memiliki kesamaan fungsi dengan yang disediakan oleh Facebook, yaitu sebagai penanda bahwa pengguna yang mengikuti suatu akun tertentu menyukai foto atau video yang telah diunggah. Berdasarkan jumlah suka pada sebuah foto atau video didalam Instagram, hal itulah yang menjadi faktor khusus yang mempengaruhi foto tersebut terkenal atau tidak.

1. Populer (Explore)

Bila sebuah foto atau video dapat memasuki kedalam halaman populer, secara tidak langsung foto atau video tersebut akan menjadi suatu hal yang dikenal oleh masyarakat mancanegara karena *explore* merupakan tempat foto atau video populer seluruh dunia dijadikan dalam satu halaman , sehingga jumlah pengikutnya juga dapat bertambah lebih banyak.

### Presentasi Diri/Personal Branding

Menurut Goffman, “Presentasi diri merupakan suatu kegiatan yang dilakukan oleh individu tertentu untuk memproduksi definisi situasi dan identitas sosial bagi para aktor dan definisi situasi tersebut mempengaruhi ragam interaksi yang layak dan tidak layak bagi para aktor dalam situasi yang ada”(Mulyana, 2003, p. 112).

### Klasifikasi

Klasifikasi adalah suatu metode pengelompokan data dimana data yang digunakan tersebut mempunyai kelas label atau target. Sehingga algoritma-algoritma untuk menyelesaikan masalah klasifikasi dikategorikan ke *dalam supervised learning* atau pembelajaran yang diawasi. Maksud dari pembelajaran yang diawasi adalah data label atau target ikut berperan sebagai supervisor atau guru yang mengawasi proses pembelajaran dalam mencapai tingkat akurasi atau presisi tertentu. Ada beberapa metode yang digunakan dalam proses klasifikasi untuk menyelesaikan suatu masalah yaitu: *backpropagation neural network, Support Vector Classification (SVC), Extreme Learning Machine (ELM), K-Nearest Neighbor, Naive Bayes*, dan lain sebagainya(Adil, 2017).

### K-Nearest Neighbor

*K-Nearest Neighbor* merupakan metode klasifikasi objek yang menggunakan pembelajaran dengan mencari jarak terdekat dari suatu objek. Data pelatihan diproyeksikan ke ruang multi-dimensi, di mana setiap dimensi mewakili karakteristik data. Ruang ini dibagi menjadi beberapa bagian sesuai dengan klasifikasi data pelatihan. *K-Nearest Neighbor* adalah metode pengklasifikasian suatu kumpulan data berdasarkan pembelajaran terhadap data yang telah diklasifikasikan sebelumnya. Termasuk dalam pembelajaran terawasi, di mana hasil *query* versi baru diurutkan sebagian besar berdasarkan kedekatan kategori di *K-Nearest Neighbor*(Fauzi, 2018).

*K-Nearest Neighbor* (KNN) juga termasuk dalam kumpulan pembelajaran berbasis instans. Algoritma ini merupakan bagian dari teknik *lazy learning*. *K-Nearest Neighbor* dilakukan dengan mencari nilai K objek pada data latih yang paling dekat (mirip) dengan objek pada data baru/tes. Algoritma *K-Nearest Neighbor* merupakan metode klasifikasi objek berdasarkan data latih yang paling dekat dengan objek tersebut. *Nearest Neighbor* adalah pendekatan untuk menemukan kasus dengan menghitung kedekatan antara kasus baru dan kasus lama berdasarkan pencocokan bobot dari beberapa fitur yang ada(Pristiyanti et al., 2018).

Terdapat beberapa cara untuk menghitung nilai jarak antar objek data pada algoritma K-Nearest Neighbor, salah satu cara yang sering digunakan adalah *Ecludiean Distance.* Berikut ini merupakan rumus dari *Ecludiean Distance* :



Gambar 2. 1 Rumus Euclidean Distance

dist(x, y) = jarak antar obyek x dan y

xi = Data Uji atau Data Testing

yi = *Data training*

n = dimensi data

### Python

Python adalah bahasa pemrograman interpretatif, berorinetasi objek dan semantik yang dinamis. Python memiliki *high-level* struktur data, *dynamic typing* dan *dynamic binding*. Python memiliki sintaks sederhana dan mudah dipelajari untuk penekanan pada kemudahan membaca dan mengurangi biaya perbaikan program. Python mendukung modul dan paket untuk mendorong kemodularan program dan *code reuse*. Interpreter Python dan standard *library*-nya tersedia secara gratis untuk semua platform dan dapat secara bebas disebarkan. Bahasa pemrograman ini dibuat oleh Guido van Rossum(Python Software Foundation, 2016.).

### Requests

Requests merupakan salah satu *library* pada Python yang berfungsi untuk melakukan pengambilan data mulai dari melempar parameter dalam URL sampai mengirim *header* khusus dan verifikasi SSL.

### Streamlit

Streamlit adalah *library framework* dari Python yang bersifat *open-source* yang dibuat untuk memudahkan membangun sebuah aplikasi *website* di bidang sains data dan *machine learning*.

### Non-Relational Database

*Non-Relational Database* atau sering disebut dengan NoSQL merupakan suatu sistem pengelolaan basis data yang tidak menggunakan manajemen relasional, atau bisa dikatakan data tidak saling mengenal relasi antar tabel dan tidak perlu menggunakan struktur yang sama pada setiap *record*nya. Terdapat beberapa tipe dari basis data ini seperti misal *key-value*, *document*, *graph*, dan *column*. Beberapa manajemen basis data yang menggunakan sistem pengelolaan ini adalah Oracle NoSQL Database, Redis, MongoDB, Virtuoso, dan Cassandra.

### Mongo Database

Mongo Database atau biasa disebut MongoDB adalah basis data yang bersifat *non-relational database* dengan memanfaatkan catatan dari dokumen yang berisi struktur data yang terdiri dari *field* dan *value pairs* dengan varian yang disebut Binary JSON(BSON). *Field* yang terdapat pada dokumen ini mirip dengan kolom pada sistem basis data relasional. Kumpulan dokumen disebut dengan *collection* yang memiliki fungsi setara dengan tabel database relasional. Sebuah *collection* dapat berisi jenis data apapun, namun data ini tidak dapat tersebar di basis data yang berbeda.