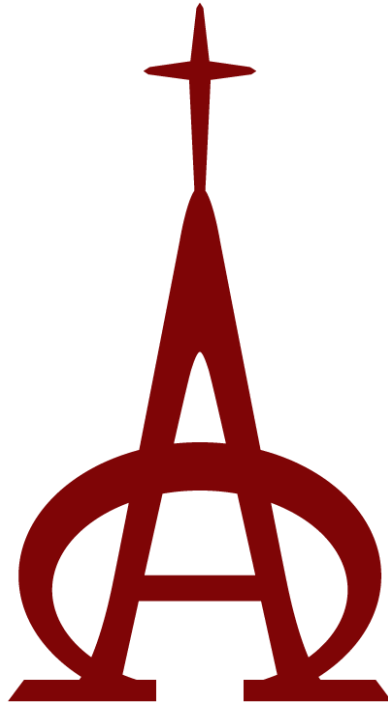


Laporan UAS IBDA 3111 Nomor 2



Ditulis oleh:

Moody Asyer - 191900154

Prodi: IBDA

CALVIN INSTITUTE OF TECHNOLOGY
JAKARTA
2021

Latar Belakang

Selama masa pandemi Covid-19 yang berlangsung hampir 2 tahun, banyak usaha mikro kecil menengah (UMKM) yang terdampak. Hanya 3.6% yang mengalami pertumbuhan sedangkan 96.4% sisanya mengalami stagnan ataupun penurunan dengan tingkat yang berbeda-beda.

Semenjak dua bulan terakhir ini jumlah kasus Covid-19 sudah menyusut, hal ini diikuti dengan menurunnya tingkat/level dari PPKM dari yang awalnya level-4 sekarang menjadi level-1. Oleh karena itu, dalam upaya mempercepat proses pemulihan dari bisnis UMKM yang terdampak, pada kasus ini yang akan dilakukan adalah membuat sebuah model pembelajaran mesin yang mampu memberikan rekomendasi lokasi yang tepat untuk membuka suatu restoran yang terkena dampak dari pandemi Covid-19, yaitu restoran padang 'Ganto Minang' di daerah Jakarta Selatan.

Model pembelajaran mesin ini nantinya akan mempelajari pola *point of interest* (POI) yang ada di sekitar restoran lalu mencari pola terbaik untuk membuka cabang dari bisnis restoran tersebut.

Paparan Studi Kasus

Pada kasus kali data yang akan digunakan adalah data dari dataset 'Tesselation.csv'. Dataset ini berisi data dari kumpulan grid lokasi yang menunjukkan informasi kategori POI dan informasi keberadaan ganto minang. nantinya pada dataset tersebut akan dilakukan sedikit pengolahan untuk mencari dan menentukan lokasi/grid yang ada restoran ganto minang dan grid sekitarnya, kemudian dilakukan teknik pra pemrosesan dengan LDA dan PCA, dan terakhir akan dibuat pelatihan dan evaluasi model pembelajaran mesin.

Analisa Hasil dari pelatihan pada file ipynb

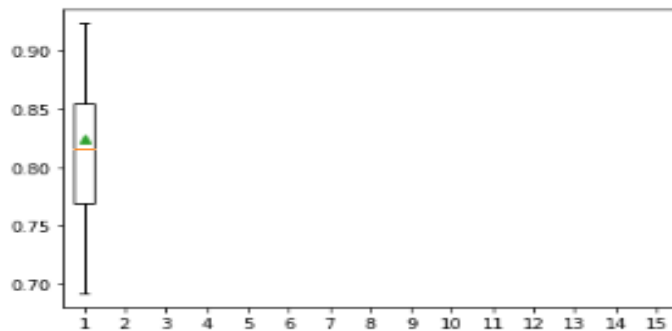
Pada kasus kali ini setelah membuat tabel/dataset baru yang kita olah sedemikian rupa sehingga hanya berisi data grid jakarta selatan yang memiliki ganto minang serta grid terdekatnya. Yang dilakukan selanjutnya adalah melakukan teknik reduksi dimensi data dengan LDA dan juga PCA. Sehingga didapat hasil sebagai berikut.

- LDA

```

>1 0.825 (0.066)
>2 nan (nan)
>3 nan (nan)
>4 nan (nan)
>5 nan (nan)
>6 nan (nan)
>7 nan (nan)
>8 nan (nan)
>9 nan (nan)
>10 nan (nan)
>11 nan (nan)
>12 nan (nan)
>13 nan (nan)
>14 nan (nan)
>15 nan (nan)

```



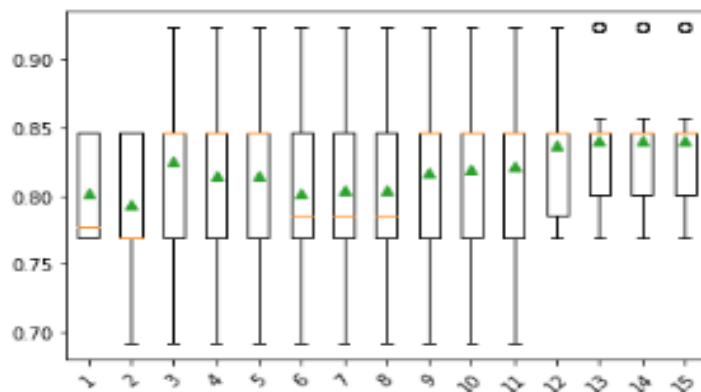
Dari hasil LDA kita melihat bahwa hanya ada 1 skor/ 1 dimensi yang didapatkan dan memiliki akurasi sebesar 82.5%. Hal ini terjadi karena pada data output/y hanya memiliki dua jenis kelas data, yaitu 'Ya' dan 'Tidak' atau dengan kata lain hanya 1 dan 0. Dengan begitu sangat wajar jika dengan menggunakan LDA pada akhirnya kita hanya mendapat hasil 1 skor/ 1 dimensi saja dan sisanya adalah nan. Karena hanya ada 1 dimensi dan dimensi tersebut memiliki akurasi tertinggi, maka nilai akurasi 82.5% inilah yang akan diambil.

- **PCA**

```

>1 0.802 (0.037)
>2 0.794 (0.044)
>3 0.825 (0.062)
>4 0.814 (0.060)
>5 0.814 (0.060)
>6 0.802 (0.060)
>7 0.804 (0.057)
>8 0.804 (0.057)
>9 0.817 (0.060)
>10 0.820 (0.060)
>11 0.822 (0.057)
>12 0.837 (0.050)
>13 0.840 (0.049)
>14 0.840 (0.049)
>15 0.840 (0.049)

```



Dari hasil PCA kita mendapatkan 15 skor/dimensi, jika kita melihat perbandingan antara 15 dimensi ini beserta dengan akurasinya maka sangat jelas bahwa setiap kali jumlah dimensi bertambah pada beberapa dimensi tertentu akan meningkat dan terkadang menurun, tetapi secara garis besar kita menyadari bahwa akurasinya meningkat.

Hal lainnya kita temui ketika kita berada di komponen ke-13/13 dimensi. Kita melihat bahwa setelah dimensi ke-13 tidak terjadi lagi perubahan akurasi (secara khusus terjadinya peningkatan). Karena itu, kita bisa mengambil keputusan bahwa 13 dimensi merupakan/memberikan akurasi yang terbaik/maximal (84%) pada kasus ini.