

BAB II KAJIAN PUSTAKA

2.1. Penelitian Terkait

Penulis memulai penelitian ini dengan terlebih dahulu melakukan studi kepustakaan dari penelitian-penelitian dan sumber-sumber lain. Dari studi kepustakaan itu, penulis menemukan beberapa penelitian yang mendorong untuk mengangkat tema seperti diatas. Adapun penelitian lain yang berkaitan dengan penelitian ini, antara lain sebagai berikut:

Penelitian yang dilakukan oleh Alena Uperiati, Martaleli Bettiza, dan Atika Puspasari dengan judul **“PERBANDINGAN METODE FUZZY C-MEANS DAN K-MEANS DALAM KLASIFIKASI KELULUSAN MAHASISWA(STUDI KASUS : JURUSAN MANAJEMEN UNIVERSITAS MARITIM RAJA ALI HAJI”** Yang menjelaskan bahwa dari hasil pengamatan di Universitas Maritim Raja Ali Haji, terdapat ketidakseimbangan dari jumlah mahasiswa yang masuk dan keluar. Mahasiswa yang masuk berjumlah besar, sedangkan jumlah mahasiswa yang lulus tepat waktu yang berdasar ketentuan jauh lebih sedikit dibandingkan dengan jumlah mahasiswa yang masuk ke Universitas Maritim Raja Ali Haji Tanjung pinang. Pada tahun 2014 mahasiswa yang masuk jauh tidak sebanding dengan mahasiswa yang lulus tepat pada waktunya. Dalam jangka waktu yang relatif panjang, hal ini dapat mengakibatkan penumpukan mahasiswa di Universitas Maritim Raja Ali Haji yang berjumlah banyak. Masalah seperti ini dapat segera diatasi dengan cara meningkatkan pembelajaran yang lebih berkualitas dan layanan akademik untuk mahasiswa. Selain itu, apabila jangka waktu penyelesaian studi mahasiswa sudah dapat diprediksi maka solusi penanganan masalah tersebut akan lebih efektif. Salah satu cara melakukan suatu prediksi yang dapat digunakan adalah dengan cara penggalian data atau data mining. Penggalian data atau data mining yang berdasarkan dari data pendidikan di perguruan tinggi dapat meningkatkan pembelajaran yang berkualitas bagi mahasiswa di perguruan tinggi. Algoritma penghitungan data mining yang digunakan didalam penelitian ini adalah algoritma *Fuzzy C-Means* dan *K-Means*. Didapatkan hasil pegujian yang dimana

perbandingan hasil tingkat akurasi antara *Fuzzy C-Means* dengan *K-Means* dipengaruhi berdasarkan inputan nilai pangkat, maksimum iterasi, dan nilai error. Sedangkan di metode *K-Means* hasil akurasinya baru didapatkan pada pengujian ini, iterasi yang sering muncul adalah iterasi ke-4. Kesimpulannya adalah dengan menggunakan metode *fuzzy c-means* didapatkan hasil akurasi sebesar 81.91%. Sedangkan dengan menggunakan metode *k-means* didapatkan hasil akurasi hanya sebesar 63%. Maka dari itu penggunaan metode *fuzzy c-means* lebih baik dari pada menggunakan metode *k-means* (Uperiati, Bettiza, & Puspasari, 2020).

Penelitian selanjutnya dilakukan oleh Faris Asrory, Joseph Dedy Irawan, dan Abdul Wahid dengan judul **“APLIKASI PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN METODE KNN** “. Yang menjelaskan bahwa perguruan tinggi diartikan sebagai suatu lembaga pendidikan yang menyediakan Pendidikan yang Tinggi. STITNU Al Mahsuni contohnya dapat di kelompokkan ke dalam Perguruan Tinggi baru apabila di bandingkan dengan perguruan tinggi lainnya. Akan tetapi STITNU Al Mahsuni memiliki visi dan misi yang kuat dan luas untuk menjadikannya Universitas atau perguruan tinggi yang besar dan dapat di perhitungkan sebagai perguruan tinggi yang berkualitas di indonesia. Demi memperoleh tujuan tersebut, di butuhkan Akreditasi yang baik dari dikti. Untuk memperoleh peningkatan akreditasi, ada beberapa syarat ketaf yang harus dipenuhi oleh pihak lembaga yang bersangkutan. Syarat utama yang paling penting adalah peningkatan kelulusan mahasiswa tepat waktu pada lembaga tersebut diharuskan lebih konsisten artinya perbandingan jumlah mahasiswa didik baru dengan mahasiswa yang lulus tepat waktu haruslah berbanding lurus atau seimbang. Akan tetapi pada kenyataannya masih ada mahasiswa yang lulus lebih dari waktu yang sudah di tentukan di STITNU Al Mahsuni. Oleh sebab itu, diperlukan suatu sistem yang dapat menemukan indikasi-indikasi terlambatnya kelulusan mahasiswa tepat waktu dan digunakanlah metode KNN. Metode *K-Nearest Neighbor* (KNN) adalah algoritma penghitungan klasifikasi ketetangaan dari jarak terdekat (neighbor) sebagai nilai hasil prediksi dari query instance yang baru. Dari hasil penggunaan metode ini didapatkan nilai akurasi tertinggi yaitu prediksi dengan menggunakan k 1,4,5,6,8,9 dan 10 yang masing-masing nilai akurasi adalah 93,33%.

Kesimpulannya adalah penggunaan metode KNN dengan parameter memiliki pekerjaan, status pernikahan, jenis kelamin, umur dan jurusan SMA yang prediksinya sudah ditemukan, maka metode ini dapat digunakan pada mahasiswa STINU Al Mahsuni Lombok Timur (Faris, Joseph., & Abdul, 2020).

Penelitian selanjutnya dilakukan oleh Rini Sovia dan Aulia Fitrul Hadi dengan judul “**MEMBANDINGKAN METODE SAW DAN MFEP DALAM PENENTUAN JURUSAN DI TINGKAT SLTA** “. Yang menjelaskan bahwa di zaman globalisasi yang sangat tinggi saat ini perkembangan teknologi informasi sangatlah pesat, dari perkembangannya yang pesat sekali masyarakat diseluruh dunia diharuskan dapat mengetahui perkembangan tersebut supaya bisa menggunakan dan memanfaatkan teknologi informasi dalam menyelesaikan suatu pekerjaan dan segala permasalahan yang dihadapinya. Oleh karena itu sistem pendukung keputusan dapat diartikan sebagai sumber kecerdasan individu yang digabung dengan kemampuan komponen-komponen bertujuan untuk memperbaiki kualitas keputusan. Sistem pendukung keputusan juga dapat diartikan sebagai sistem informasi berbasis komputer untuk manajemen pengambilan keputusan yang meyelesaikan masalah semi terstruktur. Metode *Simple Addictive Weighting* (SAW) dan *Multifactor Evaluation Process*(MFEP) adalah metode yang digunakan dalam penelitian ini. Dengan menggunakan kedua metode dalam penelitian tersebut akan menghasilkan persentase prediksi keakuratan dalam mendukung suatu keputusan untuk menentukan jurusan yang cocok dengan siswa yang berada di SMA Negeri 10 Padang. Dari hasil perhitungan V_i diperoleh hasil akhir yang dimana masing-masing siswa yang mendapatkan nilai total diatas dari $>3,00$ maka siswa tersebut masuk jurusan IPA, sedangkan bagi siswa yang memperoleh nilai total <3.00 maka siswa tersebut masuk jurusan IPS. Kesimpulannya adalah dari hasil ini didapatkan bahwa metode SAW mempunyai nilai akurasi yang lebih baik dari metode MFEP. Terbukti dari tahap standarisasi di SAW dan bukan dari MFEP. Hasil pengambilan keputusan dari kedua metode ini ditentukan sesuai dengan kriteria yang dibutuhkan oleh masing-masing metode. Ternyata SAW membutuhkan lebih banyak standar daripada MFEP. Fakta membuktikan bahwa hasil pengambilan keputusan SAW lebih baik. Dengan menggunakan Simple Addiction Weighting (SAW), akurasi analisis

menggunakan skor bahasa Indonesia, Inggris, matematika, IPA, dan psikotes adalah 38,3%, dan MFEP dengan bobot pendukung yaitu IPA, skor tes psikologi dan minat, menghasilkan akurasi 70,5% (Rini & Aulia, 2019).

Dari paparan perbandingan ke tiga jurnal penelitian di atas dengan penelitian selanjutnya adalah didapatkan persamaan menggunakan metode yang sama tapi data yang digunakan berbeda. Sehingga mendapatkan informasi yang berbeda. Maka, dari penelitian sebelumnya dilakukan penelitian yang sama dengan judul “Penerapan Data Mining Dalam Penentuan Jurusan Di MAN 1 Probolinggo Menggunakan Metode K-Nearest Neighbor”. Akan diterapkan metode KNN dimana metode ini menggunakan klasifikasi ketetangaan.

2.2. Landasan Teori

Pada landasan teori menjelaskan beberapa pengertian dan metode yang ada di penelitian ini, antara lain Data Mining, Klasifikasi, K-Nearest Neighbor, Penjurusan, Performa, dan Python.

2.2.1 Data Mining

Data Mining ialah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis (Syahra, 2018). Data Mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya

Tahapan data mining adalah sebagai berikut:

a. Pembersihan Data

Pembersihan data adalah proses menghilangkan noise dan data yang tidak konsisten atau tidak relevan.

b. Integrasi data (Data Integration)

Integrasi data adalah menggabungkan data dari berbagai database ke dalam database baru.

- c. Pemilihan data (Data Selection) Data dalam database seringkali tidak digunakan semua, sehingga hanya data yang sesuai untuk analisis yang akan diambil dari database.
- d. Transformasi Data Data diubah atau digabungkan menjadi format yang sesuai untuk diproses dalam data mining.
- e. Proses penambangan adalah proses utama penerapan metode untuk menemukan pengetahuan yang berharga dan tersembunyi dari data. Beberapa metode dapat digunakan berdasarkan pengelompokan data mining.
- f. Evaluasi pola (Pattern Evaluation) untuk mengidentifikasi pola-pola menarik dalam pengetahuan berdasarkan penemuan.
- g. Presentasi Pengetahuan (Knowledge presentation) didefinisikan sebagai visualisasi dan penyajian pengetahuan menyangkut metode yang dipakai bertujuan agar mendapatkan pengetahuan yang diperoleh pengguna.

2.2.2. Klasifikasi

Klasifikasi adalah suatu cara untuk menemukan suatu kumpulan model (fungsi) yang dapat menjelaskan dan memisahkan kelas data atau konsep yang bertujuan agar bisa memakai model tersebut untuk memperkirakan kelas dari suatu objek ke objek lainnya yang dimana kelasnya belum diketahui (Rani, 2015).

Proses klasifikasi biasanya dibagi menjadi dua tahap: pembelajaran dan pengujian. Pada fase pembelajaran, kelas data yang diketahui diumpangkan untuk membentuk model prediktif. Karena menggunakan data yang sebelumnya diberi label oleh para ahli di bidangnya sebagai contoh data yang benar, klasifikasi sering juga disebut sebagai metode terawasi. Kemudian pada tahap pengujian, model yang sudah terbentuk diuji dengan beberapa data lain untuk mengetahui keakuratan model. Jika akurasi termasuk dalam model ini, maka dapat digunakan untuk memprediksi kelas data yang tidak diketahui.

Klasifikasi adalah tindakan untuk memberikan kelompok pada setiap keadaan. Setiap keadaan berisi sekelompok atribut, salah satunya adalah *class attribute*. Metode ini butuh untuk menemukan sebuah model yang dapat menjelaskan *class attribute* itu sebagai fungsi dari input *attribute*. Proses

klasifikasi diawali dengan *preprocessing* data untuk melakukan penghilangan *missing value* dan pemilihan fitur pada dataset.

2.2.3. K-Nearest Neighbor

Menurut Susanto (Susanto, 2020) klasifikasi ketetanggaan pada metode KNN didasarkan pada jarak antara data baru dan data model. K-Nearest Neighbor ialah suatu metode yang dipakai untuk menangani masalah pengklasifikasian yaitu kelas didasarkan pada kesamaan maksimum antara data baru dan data lama atau jarak antara beberapa tetangga dengan ketidaksamaan terkecil. Tetangga terdekat ialah objek training yang mempunyai nilai persamaan terbesar atau ketidaksamaan terkecil dari data lama. Seluruh total tetangga terdekat diterangkan dengan k . Nilai k terbaik tergantung pada data.

Secara umum, nilai yang tinggi akan mengurangi dampak kebisingan pada klasifikasi, tetapi akan membuat batas antara setiap kelas lebih kabur. Secara khusus, klasifikasi diprediksi berdasarkan data sampel terdekat, yaitu lebih mudah menentukan proses jurusan siswa sesuai standar yang telah ditentukan, dan lebih mudah bagi sekolah untuk mengoptimalkan proses pengajaran sesuai jurusan yang siswa.

Pendekatan sederhana untuk menentukan nilai k yaitu :

$k = 1$ ialah tetangga terdekat

$k = \sqrt{n}$, n ialah total dari semua sampel data yang ada

Contoh didapatkan 30 sampel data, untuk menetapkan nilai k diperlukan rumus.

$$k = \sqrt{n}, = \sqrt{30}$$

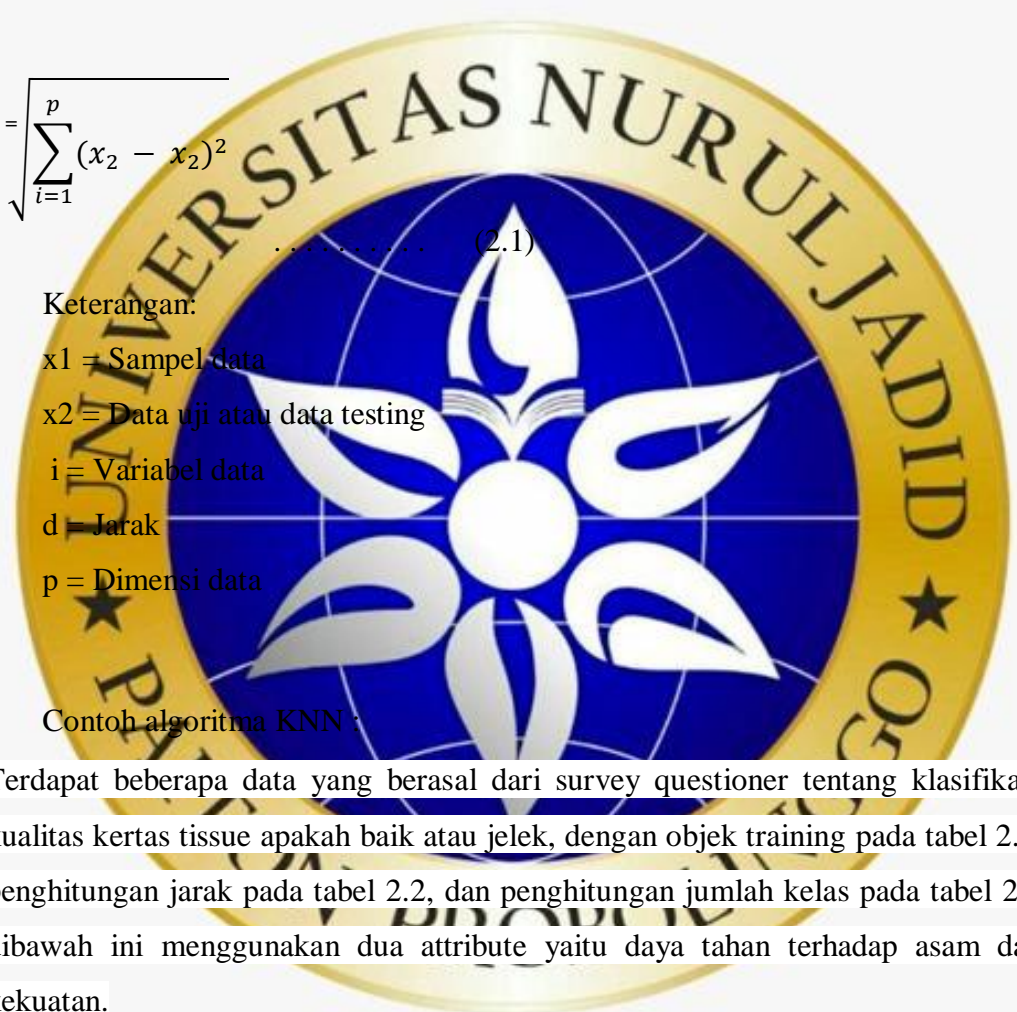
$$= 5,47 \text{ nilai } k = 5$$

Ada beberapa hal untuk menghitung metode *k-Nearest Neighbor* sebagai berikut:

1. Menetapkan parameter k .
2. Mengukur jarak antara data uji dengan semua data latih.

3. Menyusun jarak yang terbentuk.
4. Menetapkan jarak terdekat sampai ke urutan k.
5. Memasangkan kelas yang bersesuaian.
6. Menemukan total kelas dari tetangga terdekat kemudian tentukan kelas tersebut sebagai kelas data yang akan diuji.

Perhitungan jarak antara data baru dengan data lama dilakukan menggunakan rumus sebagai berikut.



$$d = \sqrt{\sum_{i=1}^p (x_1 - x_2)^2} \dots \dots \dots (2.1)$$

Keterangan:

- x1 = Sampel data
- x2 = Data uji atau data testing
- i = Variabel data
- d = Jarak
- p = Dimensi data

Contoh algoritma KNN :

Terdapat beberapa data yang berasal dari survey questioner tentang klasifikasi kualitas kertas tissue apakah baik atau jelek, dengan objek training pada tabel 2.1, penghitungan jarak pada tabel 2.2, dan penghitungan jumlah kelas pada tabel 2.3 dibawah ini menggunakan dua attribute yaitu daya tahan terhadap asam dan kekuatan.

Ditentukan sampel data = 6, $\sqrt{6} = 2,44$, berarti parameter k = 3

Tabel 2.1 Objek Training

| X1 = Daya tahan asam(detik) | X2 = Kekuatan(kg/m ²) | Y = Klasifikasi |
|-----------------------------|-----------------------------------|-----------------|
| 8 | 4 | Baik |
| 4 | 6 | Jelek |
| 7 | 7 | Baik |
| 5 | 6 | Jelek |
| 6 | 5 | Baik |

Akan diproduksi kembali kertas tissue dengan attribute **X1=7** dan **X2=4**, tanpa harus mengeluarkan biaya untuk melakukan survey, maka dapat diklasifikasikan kertas tissue tersebut termasuk yang baik atau jelek.

Tabel 2.2 Penghitungan Jarak

| X1 = Daya tahan asam(detik) | X2 = Kekuatan(kg/m ²) | Square distance to query distance (7,4) |
|-----------------------------|-----------------------------------|---|
| 8 | 4 | $(8-7)^2 + (4-4)^2 = 1$ |
| 4 | 5 | $(4-7)^2 + (5-4)^2 = 10$ |
| 4 | 6 | $(4-7)^2 + (6-4)^2 = 13$ |
| 7 | 7 | $(7-7)^2 + (7-4)^2 = 9$ |
| 5 | 6 | $(5-7)^2 + (6-4)^2 = 8$ |
| 6 | 5 | $(6-7)^2 + (5-4)^2 = 2$ |

Tabel 2.3 Penghitungan Jumlah Kelas

| X1 = Daya tahan asam(detik) | X2 = Kekuatan(kg/m ²) | Jarak Terkecil | Y = Klasifikasi |
|-----------------------------|-----------------------------------|----------------|-----------------|
| 8 | 4 | 1 | Baik |
| 6 | 5 | 2 | Baik |
| 5 | 6 | 3 | Jelek |
| 7 | 7 | 4 | Baik |
| 4 | 5 | 5 | Jelek |
| 4 | 6 | 6 | Jelek |

Dengan mengurutkan jarak terkecil, diambil K=3, maka perbandingannya adalah 2 (Baik) >1 (Jelek). Maka dapat disimpulkan kertas tissue dengan attribute **X1=7** dan **X2=4** masuk ke kelas **Baik**.

2.2.4 Performa

Kinerja klasifikasi dapat dievaluasi dengan menghitung nilai akurasi, presisi, recall dan fmeasure. Akurasi didefinisikan sebagai kedekatan antara nilai prediksi dan nilai sebenarnya. Rumus akurasi ditunjukkan pada rumus 2.2. Akurasi didefinisikan sebagai rasio item terkait yang dipilih untuk semua item yang dipilih. Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dan jawaban atas permintaan tersebut.

$$AKURASI = \frac{(TP+TN)}{(TP+TN+FP+FN)} \dots\dots\dots (2.2)$$

$$PRESISI = \frac{TP}{TP + FP} \dots\dots\dots (2.3)$$

$$RECALL = \frac{TP}{TP + FN} \dots\dots\dots (2.4)$$

$$F - Measure = 2 \frac{Presisi \times Recall}{Presisi + Recall} \dots\dots\dots (2.5)$$

Dijelaskan TP adalah True Positif , TN adalah True Negative, FP adalah False Positif dan FN adalah False Negative. TP benar positif, TN benar negatif, FP positif salah, FN salah negatif

2.2.5 Penjurusan

Penjurusan adalah proses pemilihan jurusan berdasarkan standar sekolah (Pratiwi, 2014). Dari definisi para ahli dapat ditarik kesimpulan tentang pilihan Definisi pilihan profesional adalah suatu keputusan, suatu proses di mana seseorang membuat pilihan dari beberapa alternatif. Dapat membantu mahasiswa menentukan jurusan yang terbaik dan berkualitas sesuai dengan kemampuan, bakat dan minatnya, sehingga dapat lebih fokus, dan sesuai dengan keinginan dan jurusan serta harapannya.

Penjurusan diperkenalkan sebagai upaya untuk lebih mengarahkan siswa berdasarkan minat dan kemampuan akademiknya. Umumnya SMA/MA di Indonesia mengadakan program pemilihan jurusan bagi kelas X untuk menganalisa jurusan yang tepat bagi siswanya di kelas XI nanti. Hal tersebut bertujuan untuk memberikan gambaran yang lebih jelas mengenai potensi yang dimiliki siswa agar dapat di maksimalkan oleh siswa yang bersangkutan.

2.2.6 Python

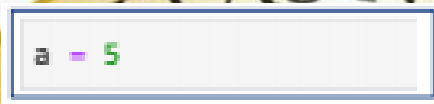
Python ialah bahasa pemrograman multiguna yang diinterpretasikan yang filosofi desainnya berfokus pada keterbacaan kode. Python dikatakan sebagai bahasa yang menggabungkan kemampuan, keahlian, dan sintaks kode yang sangat jelas, serta dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Python pula didukung oleh komunitas yang besar.

Python mendukung beberapa paradigma pemrograman penting, tetapi tidak terbatas pada pemrograman berorientasi objek, imperatif, dan fungsional. Salah satu fitur yang ada di Python adalah bahasa pemrograman dinamis dengan manajemen memori otomatis. Seperti bahasa pemrograman dinamis lainnya, Python sering digunakan sebagai bahasa skrip, tetapi dalam praktiknya penggunaannya mencakup cakupan penggunaan yang lebih luas daripada bahasa skrip yang biasanya tidak dicoba. Python dapat digunakan untuk banyak tujuan

pengembangan perangkat lunak dan dapat berjalan di banyak platform sistem operasi yang berbeda (Ratna, 2020).

1. Variabel

Variabel adalah tempat penampungan atau penyimpanan data atau nilai dengan tipe tertentu dan nilainya bisa berubah sewaktu-waktu sesuai dengan tipenya, dalam pemrograman python tidak ada cara khusus dalam mendeklarasikan variabel, artinya variabel langsung dikenali oleh python tanpa harus dideklarasikan sebelumnya. Contoh pada gambar 2.1.



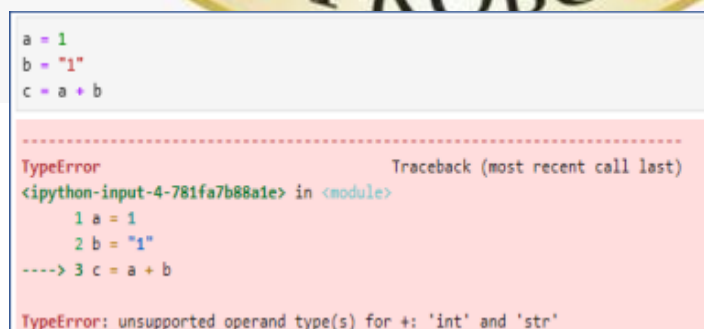
```
a = 5
```

Gambar 2.1 Contoh penulisan variabel.

Pada gambar 2.1 menunjukkan bahwa variabel “a” memiliki nilai 5. Variabel dalam python memiliki sifat fleksibel dalam mendeklarasikan variabelnya, artinya satu variabel bisa memiliki nilai yang dapat berubah-ubah dengan tipe data yang berubah juga meskipun berada dalam satu script.

2. Tipe Data

Meskipun dalam python tidak mengenal deklarasi khusus dalam menentukan tipe data, tapi python sangat memperhatikan tipe data. Contoh pada gambar 2.2.



```
a = 1
b = "1"
c = a + b
```

```
-----
TypeError                                 Traceback (most recent call last)
<ipython-input-4-781fa7b88a1e> in <module>
      1 a = 1
      2 b = "1"
----> 3 c = a + b

TypeError: unsupported operand type(s) for +: 'int' and 'str'
```

Gambar 2.2 Variabel yang memiliki nilai dengan tipe data berbeda.

Pada gambar 2.2 menunjukkan bahwa variabel yang memiliki nilai berupa bilangan tapi beda tipe, jika dikenai operasi bilangan penjumlahan maka akan menyebabkan error, hal itu terjadi karena python menganggap nilai dari kedua variabel tersebut tidak sama-sama bilangan. Di dalam python terdapat empat tipe data statis yaitu: integer, floating, string dan boolean. Deskripsi terdapat pada tabel 2.4.

Tabel 2.4 Tipe data statis pada python

| NO | Tipe Data | Contoh |
|----|-----------|--|
| 1. | Integer | -3, -2, -1, 0, 1, 2, 3 |
| 2. | Floating | 2.75, 0.25, 1.5768 |
| 3. | String | 'Python', 'Dimayah Sofiya Anjani' |
| 4. | Boolean | (True, False), (Benar, Salah), (On, Off) |

3. Print

Function `print()` digunakan untuk menampilkan ke layar. Function `print()` dapat mencetak string yang diinginkan pada layar. Pada contoh dibawah ini terdapat beberapa cara penggunaan function `print()`. Pada contoh gambar 2.3 tersebut nampak bahwa: pada baris ke-31 function `print()` digunakan untuk mencetak nilai dari variabel "nama". Pada baris ke-33 function `print()` digunakan untuk menampilkan langsung nilai dari string. Pada baris ke-34 dan baris ke-35 function `print()` digunakan untuk mencetak hasil dari operasi nilai secara langsung sedangkan pada baris ke-36 function `print()` digunakan untuk menampilkan nilai dari variabel yang digabungkan dengan string lain.

```
[31]: nama = "Dinayah Sofiya Anjani"
      print (nama)
      Dinayah Sofiya Anjani

[33]: print("Dinayah Sofiya Anjani")
      Dinayah Sofiya Anjani

[34]: print(5 + 4)
      9

[35]: print((5 + 4) * 2)
      18

[36]: nama = "Dinayah Sofiya Anjani"
      print ("Nama Saya adalah" + nama)
      Nama Saya adalahDinayah Sofiya Anjani
```

Gambar 2.3 Contoh penggunaan function print pada python.

4. Input

Function input() digunakan untuk membaca nilai yang diinput oleh user pada keyboard. Semua nilai yang dibaca dengan function input() akan dianggap string. Pada gambar 2.4 terdapat variabel nama yang akan membaca nilai input dari keyboard. Hasil pembacaan disimpan dalam variabel nama. Function print() digunakan untuk menampilkan string "Nama Saya adalah=" digabungkan dengan nilai dari variabel nama.

```
nama = input()
print("Nama Saya adalah- "+ nama)
Dinayah Sofiya Anjani

nama = input()
print("Nama Saya adalah- "+ nama)
Dinayah Sofiya Anjani
Nama Saya adalah- Dinayah Sofiya Anjani
```

Gambar 2.4 Contoh penggunaan function input pada python.

2.2.7 Google Interactive Notebook (Google Colab)

Google Colaboratory atau disebut juga *Colab* adalah tools baru yang dikeluarkan oleh [Google Internal Research](#) yang dibuat untuk membantu para Researcher dalam mengolah data untuk keperluan belajar maupun bereksperimen pada pengolahan data khususnya bidang *Machine Learning* (Digimi,2018). Pesatnya pertumbuhan bahasa pemrograman Python telah mendorong Google untuk mengembangkan program pengembangan lingkungan terintegrasi (IDE) online, yang biasa disebut sebagai *Google Interactive Notebook* atau *Google Colab*. Jenis lingkungan yang digunakan adalah *Notebook Jupyter* dengan ekstensi file * .ipynb. Sebagai referensi, Python memiliki berbagai lingkungan pemrograman, dari IDLE lama hingga Spyder penuh. Namun untuk karakter berbasis web, Google lebih memilih Notebook Jupiter.

Di sisi hardware, Google Colab menyediakan fitur seperti dukungan storage yang terintegrasi dengan Google Drive dan prosesor berupa CPU, GPU, TPU dan RAM. Selama koneksi internet lancar, sebagian besar pemrosesan dengan Google Colab baik-baik saja karena fungsionalitas server terjamin. Penelitian ini berfokus pada GPU yang saat ini banyak digunakan sebagai operator DL. GPU mengimplementasikan pemrosesan paralel yang mendukung pemrosesan data menggunakan DL dengan struktur multi-inti

Pada gambar 2.5 GPU mempunyai banyak inti untuk melakukan pemrosesan paralel.



Gambar 2.5 CPU (kiri) dengan beberapa inti sementara GPU (kanan) memiliki ratusan inti