## **2.1 *Collect Initial Data***

Hal pertama yang dilakukan pada tahap *data understanding* adalah pengumpulan data yang merupakan langkah persiapan untuk menemukan data awal. Data berasal dari beberapa sumber yang relevan dengan tujuan proyek. *Dataset* yang akan digunakan untuk memprediksi kinerja dari *employee* adalah data *employee* yang dapat diakses pada link https://www.kaggle.com/shivan118/hranalysis?select=train.csv. *Dataset* tersebut memiliki format file CSV (*Comma Separated Values*) sehingga datanya bersifat statis.

## **2.2 *Describe Data***

*Dataset* yang digunakan untuk memprediksi kinerja dari karyawan menggunakan algoritma apriori adalah predicting-employee-performance *dataset*. *Dataset* ini berisi 14 atribut atau variabel dan memuat 54808 *record*. *Dataset* yang akan dianalisis pada proyek ini terdiri dari 8 atribut nominal, 8 atribut ordinal, 3 atribut *binary*, dan 9 atribut numerik. Berikut tabel untuk menjelaskan setiap atributnya.

**Tabel 2. Tabel Atribut pada Dataset**

| No. | Nama atribut (variabel) | Tipe atribut | Deskripsi |
| --- | --- | --- | --- |
|  | *employee\_id* | Nominal | ID dari dari setiap karyawan (unik) |
|  | *department* | Nominal | Nama departemen karyawan |
|  | *region* | Nominal | Asal wilayah kerja |
|  | *education* | Nominal | Tingkat pendidikan |
|  | *gender* | Binary | Jenis kelamin karyawan |
|  | *recruitment\_channel* | Nominal | Kanal rekrutmen untuk karyawan |
|  | *no\_of\_trainings* | Numerik | Total jumlah pelatihan yang diselesaikan pada tahun sebelumnya (seperti *soft kill*, *technical skills*, dll) |
|  | *age* | Nominal | Usia karyawan |
|  | *rating* | Ordinal | Peringkat karyawan |
|  | *length\_of\_service* | Numerik | Lama bekerja dalam beberapa tahun |
|  | *KPIs\_met >8No%* | Binary | Jika persen KPI (Key Performance Indicators) lebih dari 80% => Yes, jika tidak => No |
|  | *awards\_won?* | Binary | Jika karyawan memenangkan penghargaan pada tahun sebelumnya => Yes, jika tidak => No |
|  | *avg\_training\_score* | Numerik | Skor rata-rata dalam evaluasi pelatihan |
|  | *is\_promoted* | Binary | Jika karyawan direkomendasikan untuk promosi => Yes, jika tidak => No |