

# Pemetaan Keuangan Daerah Provinsi di Indonesia Berdasarkan Realisasi Pajak Daerah dan Jumlah Populasi Pada Pemerintah Provinsi

Nur Aini Rakhmawati\*, Octgi Ristya Perdana \*, Angga Kusuma Yoga \*

Department of Information Systems, Faculty of Information Technology,

Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

nur.aini@is.its.ac.id, octgi.perdana13@mhs.is.its.ac.id, aka.yogaa@gmail.com

## Abstract

Berdasarkan data yang terdapat pada data.go.id mengenai APBD anggaran pajak daerah pada setiap provinsi yang ada pada indonesia masih belum terdapat kontrol mengenai seberapa besar selisih pendapatan dan pembelian pada masing masing provinsi yang ada pada indonesia, selain hal tersebut pemerintah Indonesia masih belum memiliki penilaian anggaran APBD pada setiap provinsi yang ada pada Indonesia, maka dari itu dengan adanya teknologi linked-data dan pengolahan data berupa klasterisasi, maka akan dilakukan penyambungan antara dataset APBD Indonesia yang terdapat pada data.go.id dengan dataset online, sehingga dapat ditarik kesimpulan mengenai apakah faktor populasi yang ada pada suatu provinsi dapat mempengaruhi pajak daerah di suatu provinsi, dan faktor apa saja yang mempengaruhi sebuah provinsi dapat bekerja secara mandiri ataupun masih bergantung terhadap pemerintah pusat, sehingga data yang telah dilakukan klasterisasi dapat menjadi tolak ukur bagi pemerintah Indonesia untuk melakukan perbaikan rencana anggaran bagi provinsi yang masih sangat bergantung terhadap pemerintah pusat.

Keywords: Pajak Daerah, Indonesia, Linked-Data, Sparql, Apache Jena Fuseki, Klasterisasi.

## 1 Pendahuluan

Dengan adanya otonomi daerah, masing-masing pemerintah daerah dapat mengelola sendiri sumber daya alam dan manusianya dengan lebih optimal, dimana otonomi daerah dimaksudkan untuk meningkatkan kemandirian daerah. Kemandirian tersebut dapat diukur berdasarkan kondisi keuangan dari daerahnya. Kondisi keuangan merupakan faktor yang penting dalam mengukur kinerja pemerintah daerah dalam mengurus rumah tangganya sendiri sehingga akan mengurangi ketergantungan pemerintah daerah terhadap bantuan dari pemerintahan pusat.

Kemandirian daerah dapat dicerminkan melalui APBD dari daerah tersebut, dimana dalam APBD tercatat semua sumber pendapatan daerah yang merepresentasikan kemampuan pemerintah dalam memanfaatkan sumber dayanya. Semakin tinggi tingkat kemandirian daerah, menandakan bahwa daerah tersebut semakin mampu memanfaatkan

sumber dayanya dan membiayai pengeluaran dari daerah tersebut tanpa tergantung terhadap bantuan pemerintah pusat.

Pendapatan dalam APBD dibagi menjadi beberapa kategori yaitu pendapatan asli daerah, dana perimbangan, dan pendapatan lainnya yang sah. Pendapatan Asli Daerah (PAD) merupakan pendapatan yang diperoleh dari sumber-sumber dalam wilayahnya sendiri seperti pajak daerah, retribusi daerah, dan hasil pengolahan kekayaan daerah. Sehingga dapat dikatakan bahwa PAD merupakan sumber pendapatan utama daerah. Selain PAD, terdapat Dana Perimbangan yang merupakan alokasi dana yang diberikan pemerintah pusat ke daerah.

Permasalahan yang timbul adalah masing-masing daerah provinsi memiliki potensi sumber daya baik alam maupun manusia yang berbeda-beda. Selain itu setiap daerah juga memiliki tingkat efisiensi dan efektivitas yang berbeda dalam memanfaatkan sumber daya tersebut. Hal ini mengakibatkan tingkat kemandirian dan juga besaran anggaran APBD yang beragam pada masing-masing daerah provinsi. Tingkat kemandirian yang tinggi mengindikasikan provinsi yang maju dan berkembang, sementara tingkat kemandirian yang rendah mengindikasikan daerah provinsi yang terbelakang. Dalam permasalahan ini, pemerintah seharusnya dapat mengambil kebijakan yang berbeda untuk setiap kelompok daerah provinsi sesuai dengan tingkat kemandirian pada masing-masing daerah.

Dengan adanya data APBD negara Indonesia yang tercatat pada data.go.id mulai tahun 2009 hingga tahun 2014, maka akan dilakukan link data pada dataset tersebut dengan beberapa dataset online yang tersedia secara universal, sehingga data tersebut dapat berguna dan dapat ditarik kesimpulan lain berdasarkan data baru yang akan ditemukan pada dataset online, seperti misalnya keterkaitan antara jumlah kepadatan penduduk pada masing-masing provinsi yang ada di Indonesia, dengan jumlah pendapatan yang dihasilkan oleh provinsi tersebut. Dari hal tersebut maka dapat berkemungkinan terjadinya anomali pada keterkaitan antara data tersebut, seperti contohnya yaitu ketika jumlah kepadatan penduduk tinggi akan tetapi pendapatan yang dihasilkan rendah. Hal tersebut dapat berguna sebagai evaluasi rencana APBD kedepannya.

Setelah dilakukan link data terhadap dataset yang ada, selanjutnya dapat dilakukan data proses, data proses yang akan dilakukan yaitu melakukan klasterisasi berdasarkan tingkat pendapatan Pajak Daerah pada masing-masing provinsi dikaitkan dengan Jumlah Populasi agar dapat terlihat pemetaan daerah berdasarkan cluster dari kondisi ekonomi masing-masing daerah. Selain itu dengan adanya data yang sangat banyak maka akan dilakukan peramalan data APBD pada setiap provinsi yang ada pada Indonesia, hal tersebut dapat berguna sebagai tolak ukur pemerintah dalam merencanakan APBD kedepannya.

Sehingga output dari apa yang akan dikerjakan yaitu dapat ditariknya kesimpulan baru mengenai data yang telah ada pada data.go.id dengan data yang tersedia secara online, selain itu juga dapat terlihat provinsi mana sajakah yang dapat mengelola APBD dengan baik dan provinsi mana yang tidak dapat mengelola APBD dengan baik, sehingga pemerintah pusat dapat berfokus untuk melakukan pengembangan kepada daerah provinsi yang masih belum dapat mengelola dana APBDnya dengan baik.

## 2 Tinjauan Pustaka

Pada bab ini akan dijelaskan penelitian sebelumnya dan dasar-dasar teori yang diharapkan dapat memberikan wawasan terkait hal-hal apa saja yang akan dilibatkan pada

penelitian guna memberikan gambaran umum pengerjaan final project ini.

## 2.1 Penelitian Terkait

Berikut merupakan penelitian yang terkait dengan project yang akan di lakukan.

| Penelitian Terkait  |  |   |
|---|--|---|
| Judul   | Penulis                                      | Keterkaitan   |
| Analisis Data Anggaran Pendapatan Belanja Daerah Menggunakan Clustering K Means dan Forecasting (Studi Kasus pada DPKA Kota Padang) | Hafilah Hamimi                               | Pada penelitian ini dilakukan pengelompokan data APBD Kota Padang pada tahun 2009 sampai 2013 menggunakan metode cluster dengan algoritma k means. Selanjutnya berdasarkan data tersebut dilakukan peramalan untuk APBD Kota Padang tahun 2014 menggunakan multiple linear regression. Kekurangan dari paper ini adalah jumlah variabel independen yang digunakan untuk clustering sedikit dimana masih banyak terdapat jenis pendapatan lainnya yang juga dimiliki oleh setiap daerah. Dalam melakukan peramalan juga tidak ditampilkan forecast error sehingga pembaca tidak dapat mengetahui seberapa akurat hasil peramalan yang telah dilakukan. |
| LinkedSpending: OpenSpending becomes Linked Open Data   | Konrad Höffner, Michael Martin, Jens Lehmann | Pada penelitian ini dilakukan konversi data transaksi keuangan pemerintah dari berbagai negara ke dalam linked data sehingga setiap data keuangan saling terintegrasi dan transparan. Keterkaitan dengan penelitian ini adalah dilakukannya linked data namun dengan fokus pada anggaran pemerintah (khususnya pendapatan) seluruh provinsi di Indonesia.   |

| Penelitian Terkait  |                     |  |
|---|---------------------|--|
| Judul   | Penulis             | Keterkaitan  |
| Analisis Pajak, Re-tribusi, DAU, BHP dan BHPB: Sebuah Pendekatan Cluster (Studi Pada 8 Pemerintah Kota dan 29 Pemerintah Kabupaten di Propinsi Jawa Timur Periode 2002 sampai 2007) | Ardi Hamzah         | Pada penelitian ini dilakukan pemetaan keuangan daerah pada 37 kabupaten atau kota di Jawa Timur menggunakan pendekatan cluster dengan menganalisis faktor (variabel independen) PAD, BHP, BHPB dan DAU. Hasilnya didapatkan tiga kluster yang disimpulkan sebagai daerah mandiri, semi mandiri, dan bergantung dimana diketahui perbedaan PAD, BHP, BHPB, dan DAU antara cluster yang satu dengan cluster yang lain. Kekurangan dari paper ini adalah jumlah variabel independen yang digunakan untuk clustering sedikit dimana masih banyak terdapat jenis pendapatan lainnya yang juga dimiliki oleh setiap daerah. |
| Pemetaan Keuangan Daerah: Pendekatan Cluster (Studi pada APBD Pemerintah Kota atau Kabupaten di Jawa Timur Tahun 2001 sampai 2006)  | Levi Olivia Santoso | Pada penelitian ini dilakukan pemetaan keuangan daerah pada 38 kabupaten atau kota di Jawa Timur menggunakan pendekatan cluster dengan menganalisis faktor (variabel independen) PAD, BHP, dan BHPB. Hasilnya didapatkan tiga kluster yang disimpulkan sebagai daerah mandiri, berkembang, dan bergantung. Kekurangan dari paper ini adalah jumlah variabel independen yang digunakan untuk clustering sedikit dimana masih banyak terdapat jenis pendapatan lainnya yang juga dimiliki oleh setiap daerah.  |

Maka dari itu, paper di atas akan menjadi acuan penelitian ini dimana dilakukan link data pada dataset tersebut dengan beberapa dataset online yang tersedia secara universal. Selanjutnya digunakannya cluster untuk mengelompokkan seluruh provinsi yang ada di Indonesia berdasarkan APBD tahun 2009-2014 dengan melibatkan lebih banyak variabel yaitu jenis-jenis pendapatan dalam APBD dan jumlah kepadatan penduduk di masing-masing provinsi.

## **2.2 Dasar Teori**

### **2.2.1 Anggaran Pendapatan, dan Belanja Daerah (APBD)**

Anggaran Pendapatan, dan Belanja Daerah (APBD) merupakan rencana keuangan yang telah dibahas dan disetujui oleh pemerintah daerah dan DPRD setempat, serta ditetapkan dengan peraturan daerah (Permendagri No.13 Tahun 2006). Salah satu anggaran dalam APBD tersebut adalah pendapatan daerah dimana menurut pasal 79 Undang-undang Nomor 22 Tahun 1999, pasal 3 dan 4 Undang-undang Nomor 25 Tahun 1999, dan pasal 157 Undang-undang Nomor 32 Tahun 2004, sumber pendapatan daerah terdiri dari:

1. Pendapatan Asli Daerah (PAD) yang terdiri dari pajak daerah, retribusi daerah, hasil pengelolaan kekayaan daerah yang dipisahkan, dan lain-lain pendapatan asli daerah yang sah.
2. Dana Perimbangan yang terdiri dari dana bagi hasil pajak, dana bagi hasil bukan pajak, dana alokasi umum (DAU), dan dana alokasi khusus (DAK).
3. Lain-lain Pendapatan Daerah yang Sah.

### **2.2.2 Link Data**

Merupakan suatu cara dalam menghubungkan data terstruktur di dalam web menggunakan standar internasional dari world wide web consortium. Tim Berners Lee sebagai penggagas linked data, memperkenalkan 4 prinsip dasar dalam melakukan linked data dari berbagai sumber data yang ada [bukunya bu iin]:

1. Menggunakan URI sebagai identitas dari sesuatu seperti orang, objek, dan lokasi.
2. Menggunakan http URI sehingga data/informasi dapat ditelusuri lebih lanjut pada web yang tersedia.
3. Informasi yang disebarkan dalam bentuk format yang standar seperti RDF.
4. Menghubungkannya dengan URI yang lain sehingga hubungan antar data dapat dilanjutkan lebih luas.

### **2.2.3 Clustering**

Metode clustering merupakan salah satu bentuk data mining yang bertujuan untuk mengelompokkan data dengan karakteristik yang sama ke dalam satu kelompok dan data dengan karakteristik yang berbeda akan masuk ke dalam kelompok yang lain. Ada beberapa pendekatan yang digunakan dalam mengembangkan metode clustering dimana salah satunya adalah partition-based clustering yaitu mengelompokkan data dengan memilah data yang dianalisis ke dalam cluster yang paling optimal.

### **2.2.4 K Means**

Salah satu algoritma yang dapat digunakan dalam partition-based clustering adalah k-means. Algoritma ini melakukan pengelompokkan data dengan prosedur sebagai berikut:

- Tentukan jumlah cluster awal.

- Alokasikan data secara random ke cluster yang ada.
- Hitung rata-rata setiap cluster dari data yang tergabung di dalamnya.
- Ulang proses nomor 3, sampai tidak ada perubahan atau perubahan yang terjadi masih di bawah threshold.

## 2.3 Geospatial

Pada bagian berikut merupakan metode yang digunakan untuk melakukan plotting pada data Pajak Daerah yang ada pada setiap provinsi di Indonesia, metode yang digunakan untuk melakukan plotting tersebut yaitu menggunakan API dari leaflet yang dapat menampilkan peta yang ada pada Indonesia.

## 3 Metodologi

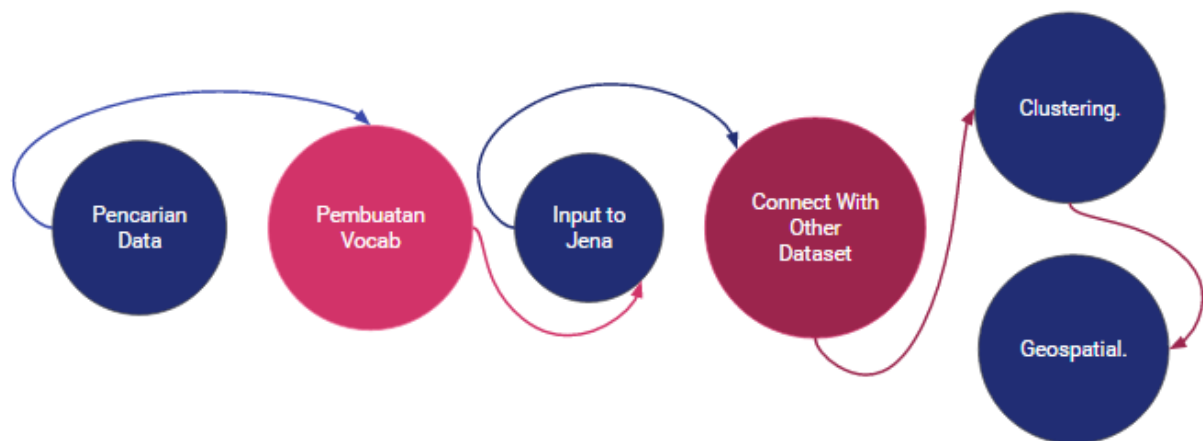


Figure 1: Metodologi Pengerjaan Project

Dalam melakukan pengerjaan project akan menggunakan metodologi sesuai pada Figure 1, berikut merupakan penjelasan dari masing masing bagian pada metodologi tersebut.

### 3.1 Pencarian Data

Pada bagian ini akan dilakukan pencarian data mengenai APBD yang direncanakan akan menggunakan data dari data.go.id, data tersebut akan digunakan sebagai dataset lokal yang akan dilakukan download secara langsung dalam bentuk csv, Figure 2 merupakan contoh dataset lokal yang akan digunakan, pada dataset yang ada pada data.go.id tersebut merupakan data APBD dari pemerintah provinsi di Indonesia.

### 3.2 Pembuatan Vocab

Pada bagian ini akan dilakukan pembuatan Vocab dari dataset lokal yang telah didapatkan, sehingga nantinya vocab dapat dilakukan query dan menampilkan data berupa linked data. Figure 3 merupakan contoh perubahan data menjadi vocab.

| _id | tahun | kode_pr... | nama_pr...   | latitude | longitude | akun       | kelompok      | jenis         | nilai_ang... |
|-----|-------|------------|--------------|----------|-----------|------------|---------------|---------------|--------------|
| 1   | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Pendapat...   | Pajak Da...   | 47700000...  |
| 2   | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Pendapat...   | Retribusi ... | 13260000...  |
| 3   | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Pendapat...   | Hasil Pen...  | 74510000...  |
| 4   | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Pendapat...   | PAD Lain ...  | 23110000...  |
| 5   | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Dana Peri...  | Bagi Hasi...  | 16500000...  |
| 6   | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Dana Peri...  | Dana Alo...   | 50970000...  |
| 7   | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Dana Peri...  | Dana Alo...   | 48190000...  |
| 8   | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Lain-lain ... | Pendapat...   | 0            |
| 9   | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Lain-lain ... | Dana Dar...   | 0            |
| 10  | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Lain-lain ... | Bagi Hasi...  | 0            |
| 11  | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Lain-lain ... | Dana Pen...   | 37280000...  |
| 12  | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Pendapatan | Lain-lain ... | Bantuan ...   | 0            |
| 13  | 2009  | 1100       | Prov. Nan... | 5.55     | 95.31667  | Belanja    | Belanja Ti... | Belanja P...  | 83480000...  |

Figure 2: Contoh dari dataset lokal mengenai APBD

### 3.3 Upload Vocabulary ke Apache Jena

Pada bagian ini data yang telah dilakukan convert menjadi vocabulary dengan bentuk turtle akan dilakukan upload pada apache jena untuk dapat dilakukan query menggunakan sparql. Query yang akan dilakukan untuk dapat menarik kesimpulan pada dataset lokal yang ada, dapat di lihat pada Figur 4.

### 3.4 Connect Dengan Dataset Lain

Pada bagian ini akan dilakukan koneksi dataset lokal dengan dataset lain yang dapat berkemungkinan lokal maupun online, dataset lain tersebut berkaitan dengan APBD akantetapi dapat memiliki kemungkinan dataset lain memiliki dimensi data lain sehingga dapat menjadi input sebagai penarikan kesimpulan yang dilakukan, pengkoneksian dengan dataset lain ini juga berguna agar dataset yang dimiliki dapat di akses oleh pengguna peneliti linked data lainnya.

### 3.5 Klasterisasi

Pada bagian ini akan dilakukan perbandingan dataset lokal1 yaitu dataset Pajak Daerah yang di dapatkan pada data.go.id dengan dataset Jumlah Populasi yang terdapat pada online seperti yang tampak pada Figure 5, sehingga akan dapat menghasilkan kesimpulan baru yang mungkin berguna untuk pemerintahan Indonesia dalam mengelola APBD untuk setiap provinsi.

### 3.6 Geospatial

Pada bagian ini akan dilakukan plotting pada data setiap provinsi yang ada di Indonesia, data hasil klasterisasi dan informasi yang ada pada setiap provinsi akan ditampilkan pada peta sehingga dapat memudahkan pengguna untuk mendapatkan informasi yang ada seperti yang tampak pada Figure 6.

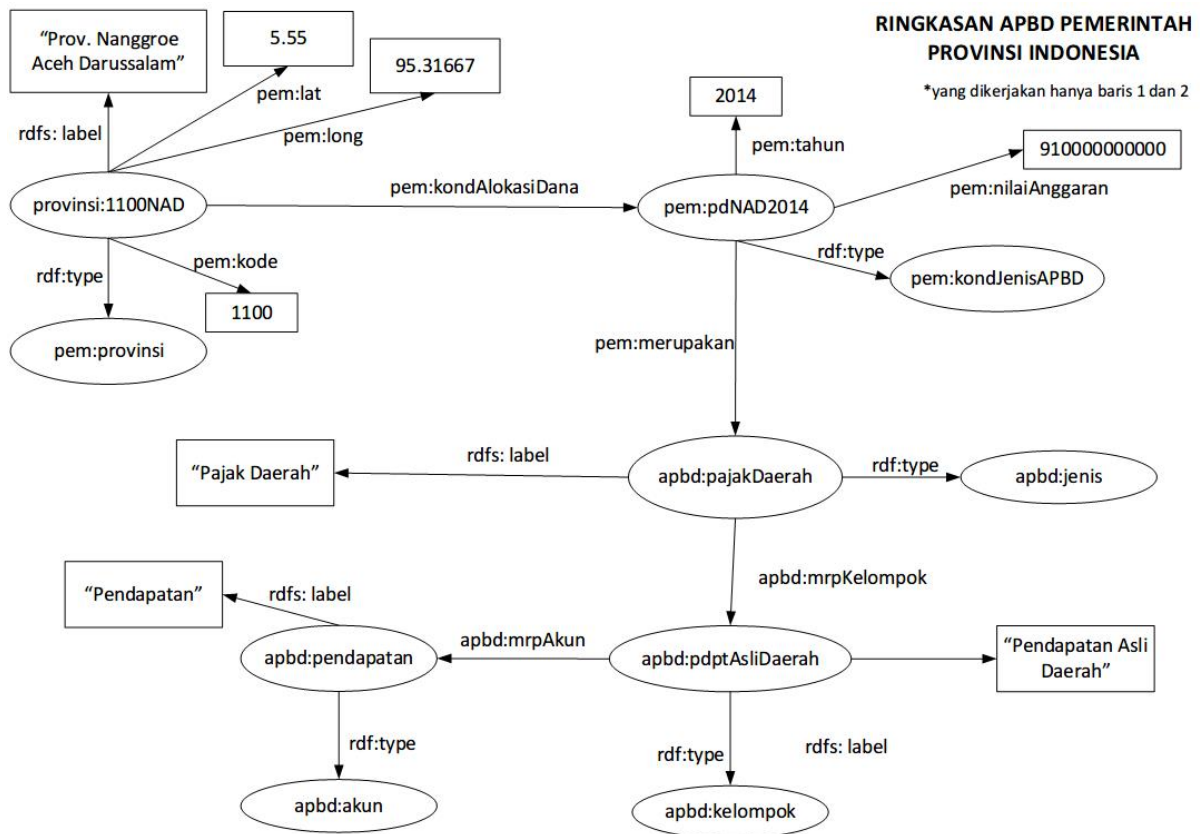


Figure 3: Contoh perubahan data menjadi vocab.

## 4 Data RDF

### 4.1 Deskripsi Dataset

Dataset lokal yang dipilih yaitu dataset mengenai APBD setiap provinsi yang ada pada Indonesia, dataset tersebut didapatkan pada website data.go.id yang menyediakan 6000 tuple pada total data, dan memiliki 10 atribut pada kolom data, data yang disediakan merupakan data APBD sejak tahun 2009 hingga 2016. Berikut merupakan Atribut yang ada pada data.

1. tahun: Tahun
2. kodeProvinsi : Kode Provinsi
3. namaProvinsi: Nama Provinsi
4. latitude: Titik Latitude Provinsi
5. longitude: Titik Longitude Provinsi
6. akun: Akun dari Anggaran
7. kelompok: Kelompok dari Akun
8. jenis: Jenis dari Kelompok
9. nilaiAnggaran: Nilai Anggaran (dalam Rupiah)



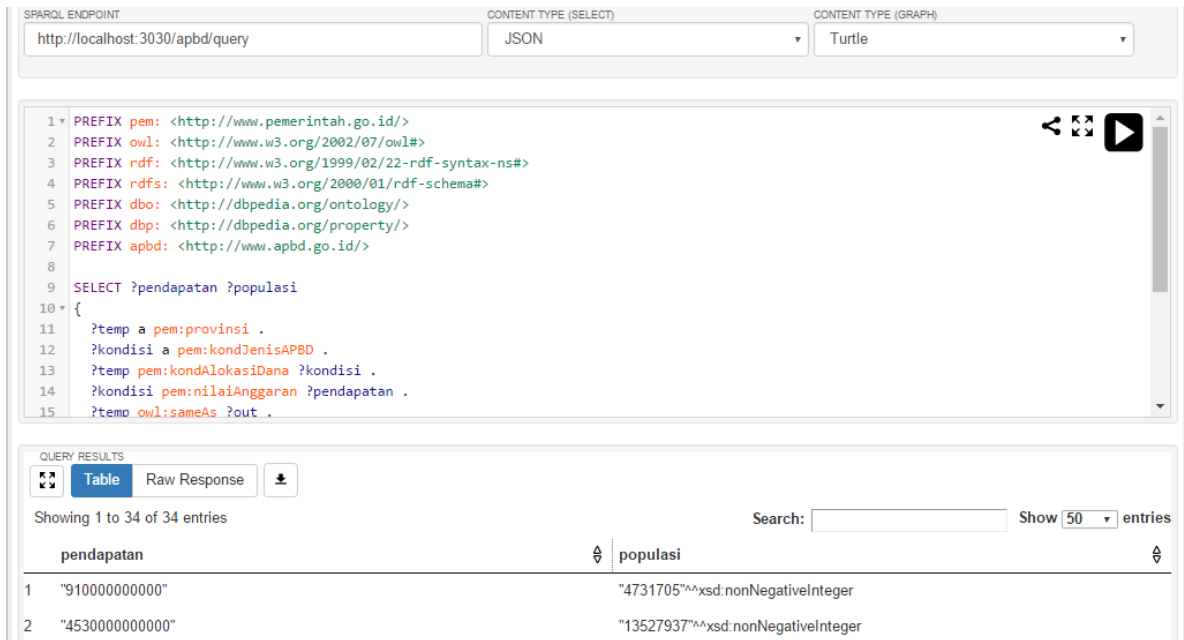


Figure 4: Melakukan Query pada Apache Jena.

## 4.2 Vocabulary

Berikut Merupakan Vocabulary dan contoh 1 instance dari vocabulary yang telah dibuat.

Listing 1: Contoh 1 SPARQL Query Mencari Atlet Indonesia di DBpedia

```

@prefix pem: <http://www.pemerintah.go.id/> .
@prefix provinsi: <http://www.provinsi.go.id/> .
@prefix apbd: <http://www.apbd.go.id/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dbr: <http://dbpedia.org/resource/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix dbo: <http://dbpedia.org/ontology/> .

pem:provinsi a rdfs:Class .
pem:kondJenisAPBD a rdfs:Class .
apbd:akun a rdfs:Class .
apbd:kelompok a rdfs:Class .
apbd:jenis a rdfs:Class .

pem:kode a rdf:Property .
pem:kode rdfs:domain pem:provinsi .
pem:kode rdfs:range xsd:integer .

pem:lat a rdf:Property .
pem:lat rdfs:domain pem:provinsi .
pem:lat rdfs:range xsd:double .

pem:long a rdf:Property .
pem:long rdfs:domain pem:provinsi .

```

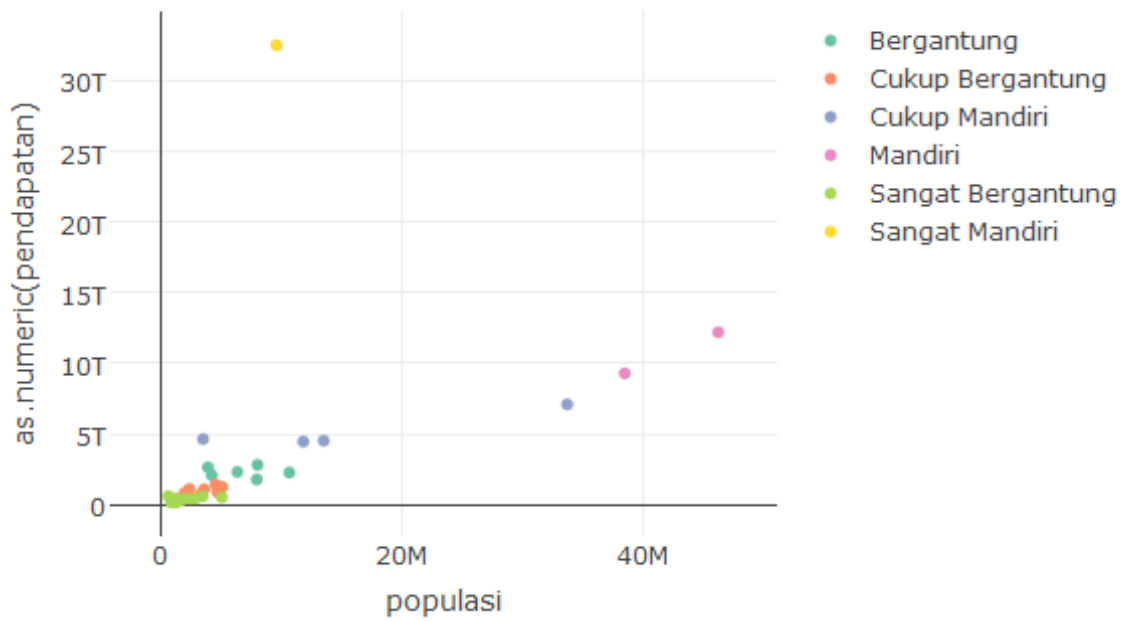


Figure 5: Klasterisasi pada Pajak setiap provinsi.

```

pem:long rdfs:range xsd:double .

pem:kondAlokasiDana a rdf:Property .
pem:kondAlokasiDana rdfs:domain pem:provinsi .
pem:kondAlokasiDana rdfs:range pem:kondJenisAPBD .

pem:nilaiAnggaran a rdf:Property .
pem:nilaiAnggaran rdfs:domain pem:kondJenisAPBD .
pem:nilaiAnggaran rdfs:range xsd:long .

pem:tahun a rdf:Property .
pem:tahun rdfs:domain pem:kondJenisAPBD .
pem:tahun rdfs:range xsd:integer .

pem:merupakan a rdf:Property .
pem:merupakan rdfs:domain pem:kondJenisAPBD .
pem:merupakan rdfs:range apbd:jenis .

apbd:mrpKelompok a rdf:Property .
apbd:mrpKelompok rdfs:domain apbd:jenis .
apbd:mrpKelompok rdfs:range apbd:kelompok .

apbd:mrpAkun a rdf:Property .
apbd:mrpAkun rdfs:domain apbd:kelompok .
apbd:mrpAkun rdfs:range apbd:akun .

provinsi:1100NAD a pem:provinsi .
provinsi:1100NAD rdfs:label "Prov. Nanggroe Aceh Darussalam" .

```

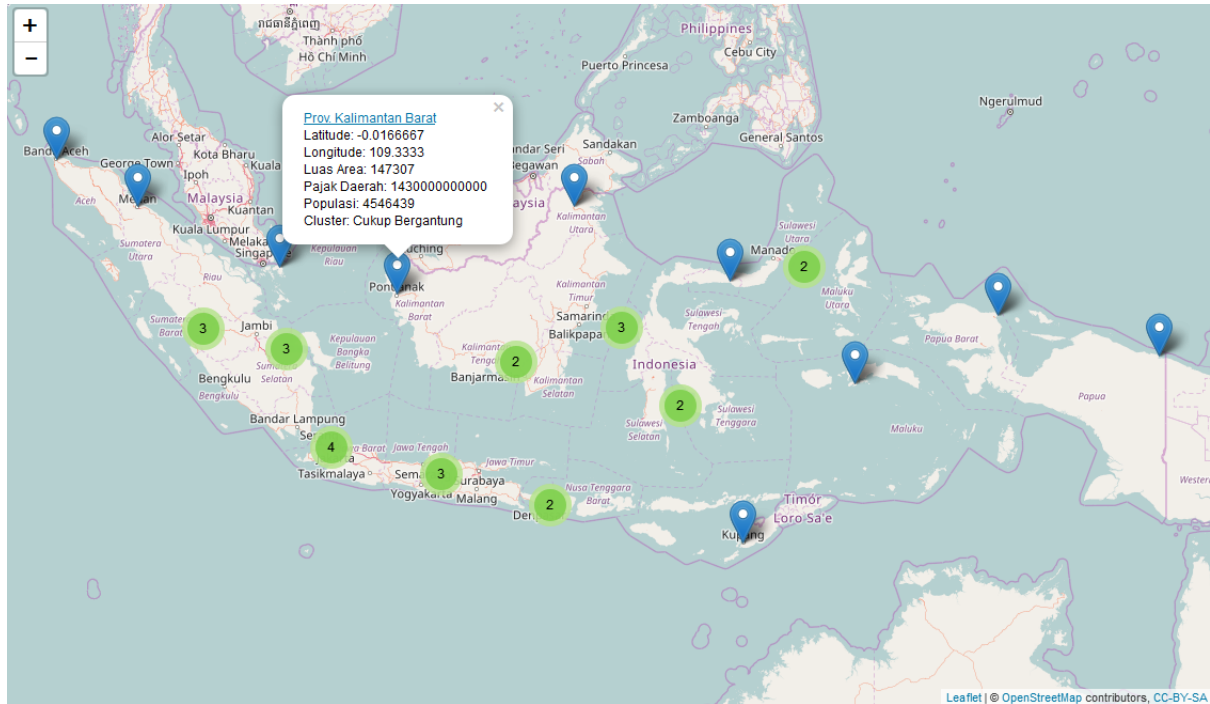


Figure 6: Pemetaan data yang telah dilakukan pengolahan.

```

provinsi:1100NAD pem:kode "1100" .
provinsi:1100NAD pem:lat "5.55" .
provinsi:1100NAD pem:long "95.31667" .
provinsi:1100NAD pem:kondAlokasiDana pem:pdNAD2014 .

pem:pdNAD2014 a pem:kondJenisAPBD .
pem:pdNAD2014 pem:tahun "2014" .
pem:pdNAD2014 pem:nilaiAnggaran "910000000000" .
pem:pdNAD2014 pem:merupakan apbd:pajakDaerah .

apbd:pajakDaerah a apbd:jenis .
apbd:pajakDaerah rdfs:label "Pajak_Daerah" .
apbd:pajakDaerah apbd:mrpKelompok apbd:pdptAsliDaerah .

apbd:pdptAsliDaerah a apbd:kelompok .
apbd:pdptAsliDaerah rdfs:label "Pendapatan_Asli_Daerah" .
apbd:pdptAsliDaerah apbd:mrpAkun apbd:pendapatan .

apbd:pendapatan a apbd:akun .
apbd:pendapatan rdfs:label "Pendapatan" .

provinsi:1100NAD owl:sameAs dbr:Aceh .
provinsi:1200Sumut owl:sameAs dbr:North_Sumatra .
provinsi:1300Sumbar owl:sameAs dbr:West_Sumatra .
provinsi:1400Riau owl:sameAs dbr:Riau .
provinsi:1500Jambi owl:sameAs dbr:Jambi .
provinsi:1600Sumsel owl:sameAs dbr:South_Sumatra .
provinsi:1700Bengkulu owl:sameAs dbr:Bengkulu .

```

```

provinsi:1800Lampung owl:sameAs dbr:Lampung .
provinsi:1900Bangka owl:sameAs dbr:Bangka–Belitung_Islands .
provinsi:2100KepRiau owl:sameAs dbr:Riau_Islands .
provinsi:3100Jakarta owl:sameAs dbr:Jakarta .
provinsi:3200Jabar owl:sameAs dbr:West_Java .
provinsi:3300Jateng owl:sameAs dbr:Central_Java .
provinsi:3400Yogya owl:sameAs dbr:Special_Region_of_Yogyakarta .
provinsi:3500Jatim owl:sameAs dbr:East_Java .
provinsi:3600Banten owl:sameAs dbr:Banten .
provinsi:5100Bali owl:sameAs dbr:Bali .
provinsi:5200NTB owl:sameAs dbr:West_Nusa_Tenggara .
provinsi:5300NTT owl:sameAs dbr:East_Nusa_Tenggara .
provinsi:6100Kalbar owl:sameAs dbr:West_Kalimantan .
provinsi:6200Kalteng owl:sameAs dbr:Central_Kalimantan .
provinsi:6300Kalsel owl:sameAs dbr:South_Kalimantan .
provinsi:6400Kaltim owl:sameAs dbr:East_Kalimantan .
provinsi:6500Kaltara owl:sameAs dbr:North_Kalimantan .
provinsi:7100Sulut owl:sameAs dbr:North_Sulawesi .
provinsi:7200Sultenga owl:sameAs dbr:Central_Sulawesi .
provinsi:7300Sulsel owl:sameAs dbr:South_Sulawesi .
provinsi:7400Sulteng owl:sameAs dbr:Southeast_Sulawesi .
provinsi:7500Gorontalo owl:sameAs dbr:Gorontalo .
provinsi:7600Sulbar owl:sameAs dbr:West_Sulawesi .
provinsi:8100Maluku owl:sameAs <http://dbpedia.org/resource/Maluku_(
    province)> .
provinsi:8200Malut owl:sameAs dbr:North_Maluku .
provinsi:9100Papbar owl:sameAs <http://dbpedia.org/resource/
    West_Papua_(province)> .
provinsi:9400Papua owl:sameAs <http://dbpedia.org/resource/Papua_(
    province)> .

```

### 4.3 Query data

Pada tahap Query data ini dilakukan 2 kali query untuk memudahkan pengolahan data, Query pertama yang di lakukan yaitu untuk mengambil variable yang akan di gunakan untuk cluster, dan query ke 2 digunakan untuk menggabungkan hasil cluster dengan seluruh data yang ada, untuk di tampilkan pada peta.

Query 1 :

Listing 2: Query untuk mengambil data yang di gunakan untuk klaster

```

PREFIX pem: <http://www.pemerintah.go.id/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbp: <http://dbpedia.org/property/>
PREFIX apbd: <http://www.apbd.go.id/>

SELECT ?pendapatan ?populasi
{

```

```

?provLokal a pem:provinsi .
?apbd a pem:kondJenisAPBD .
?provLokal pem:kondAlokasiDana ?apbd .
?apbd pem:nilaiAnggaran ?pendapatan .
?provLokal owl:sameAs ?provOnline .
SERVICE <http://dbpedia.org/sparql>
{SELECT * WHERE{
?provOnline rdfs:label ?provinsi .
?provOnline dbo:populationTotal ?populasi .
FILTER(LANG(?provinsi) = '' || LANGMATCHES(LANG(?provinsi), 'en'))
}
}
}

```

Query 2 :

Listing 3: Query yang digunakan untuk mengambil seluruh data informasi yang ada pada setiap provinsi.

```

PREFIX pem: <http://www.pemerintah.go.id/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbp: <http://dbpedia.org/property/>
PREFIX apbd: <http://www.apbd.go.id/>

SELECT ?pendapatan ?populasi
{
?provLokal a pem:provinsi .
?apbd a pem:kondJenisAPBD .
?provLokal pem:kondAlokasiDana ?apbd .
?apbd pem:nilaiAnggaran ?pendapatan .
?provLokal owl:sameAs ?provOnline .
SERVICE <http://dbpedia.org/sparql>
{SELECT * WHERE{
?provOnline rdfs:label ?provinsi .
?provOnline dbo:populationTotal ?populasi .
FILTER(LANG(?provinsi) = '' || LANGMATCHES(LANG(?provinsi), 'en'))
}
}
}

```

## 5 Analisa Data

Bada sub bagian ini akan dilakukan analisa pada data, analisa pada data yang dilakukan yaitu klasterisasi menggunakan algoritma K-Means.

## 5.1 Algoritma

Pada bagian ini akan dijelaskan tentang algoritma yang digunakan untuk melakukan klasifikasi pada data, berikut merupakan kode yang digunakan untuk klasterisasi pada data.

Listing 4: Query yang digunakan untuk mengambil seluruh data informasi yang ada pada setiap provinsi.

```
#jika di butuhkan
#install.packages("SPARQL")
#install.packages("plotly")
#install.packages("leaflet")

library(SPARQL)
library(plotly)
library(leaflet)

endpoint <- "http://localhost:3030/final-project/sparql"

#Memunculkan Variabel Pendapatan (Pajak Daerah) dan Populasi
query1 <-
"PREFIX_pem:<http://www.pemerintah.go.id/>
PREFIX_owl:<http://www.w3.org/2002/07/owl#>
PREFIX_rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX_rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX_dbo:<http://dbpedia.org/ontology/>
PREFIX_dbp:<http://dbpedia.org/property/>
PREFIX_apbd:<http://www.apbd.go.id/>

SELECT_?pendapatan_?populasi
{
  _?provLokal_a_pem:provinsi_.
  _?apbd_a_pem:kondJenisAPBD_.
  _?provLokal_pem:kondAlokasiDana_?apbd_.
  _?apbd_pem:nilaiAnggaran_?pendapatan_.
  _?provLokal_owl:sameAs_?provOnline_.
  _SERVICE<http://dbpedia.org/sparql>
  _{SELECT_*_WHERE{
    _?provOnline_rdfs:label_?provinsi_.
    _?provOnline_dbo:populationTotal_?populasi_.
    _FILTER(LANG(?provinsi) =_''_|_!LANGMATCHES(LANG(?provinsi),_'en'))
  }
}
}"

connect1 <- SPARQL(endpoint, query1)
dataframe1 <- connect1$results

summary(dataframe1)
print(dataframe1)

#Melakukan Klustering Menggunakan K-Means
result1 <- kmeans(dataframe1, 6)
```

```

result1

#Merge dataframe1 dengan Hasil result1
result2 <- data.frame(dataframe1, result1$cluster)
result2

```

Pada kode diatas dituliskan bahwa query yang digunakan untuk mengambil data kemudian disimpan pada variable query1, setelah data dimasukkan kedalam R sesuai dengan endpoint yang ada dan disimpan pada variable dataframe1, kemudian data yang ada pada dataframe1 dilakukan klasterisasi dengan 6 cluster, mengapa 6 cluster, dikarenakan 6 cluster merupakan hasil yang terbaik untuk melakukan klasterisasi dengan K-Means.

## 5.2 Hasil analisa

Setelah dilakukan proses pada data kemudian data yang telah dilakukan penampilan menggunakan grafik dan menggunakan peta, penampilan pada grafik bertujuan untuk memperlihatkan plotting dari cluster data, selain itu juga terdapat penampilan data menggunakan peta, hal tersebut bertujuan untuk memudahkan pengguna untuk mengetahui informasi setiap provinsi. Pada Figure 7 terlihat bahwa data telah dilakukan plotting menggunakan plotly pada plotting yang ada telah ditentukan berdasarkan range terbaik hingga terburuk.

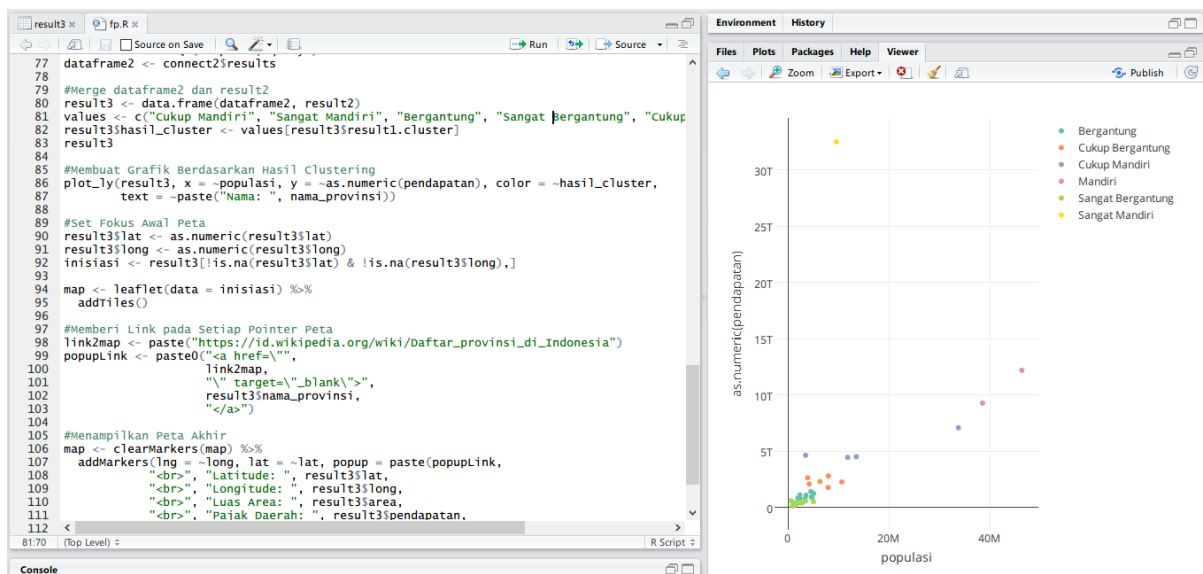


Figure 7: Melakukan plotting data pada plotly.

Setelah melakukan plotting dalam bentuk grafik selanjutnya plotting dilakukan dalam bentuk peta, Figur 8 menampilkan grafik dalam bentuk peta, dengan menggunakan library liaflet.

## 6 Kesimpulan

Berdasarkan hasil dan pembahasan tersebut, maka kesimpulan dari penelitian ini adalah sebagai berikut:

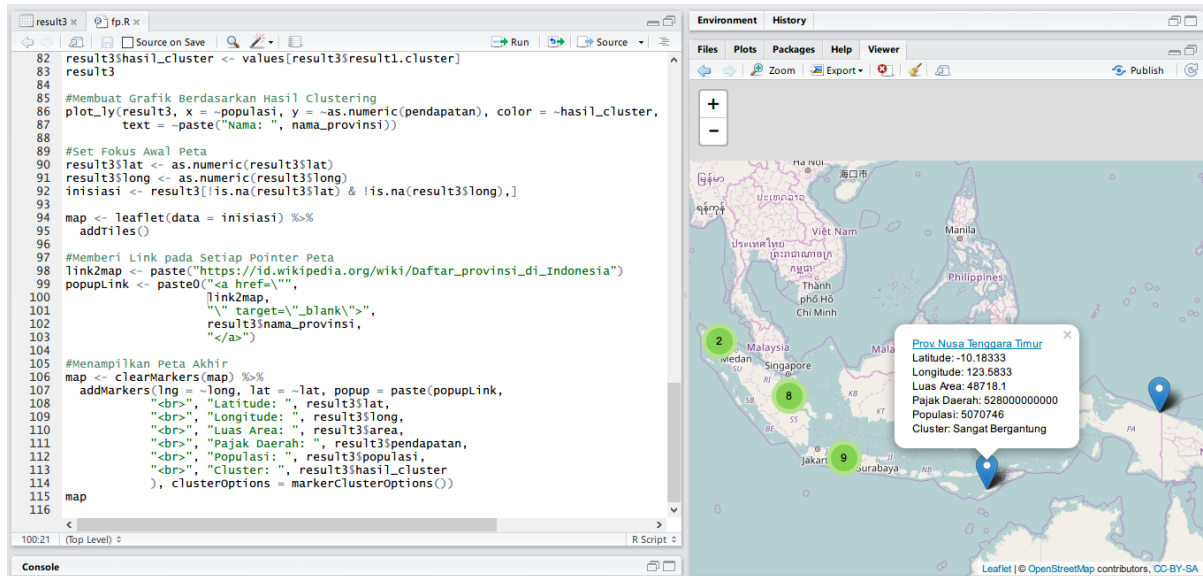


Figure 8: Melakukan plotting data pada peta.

- Terbentuk enam cluster yang masing-masing cluster mempunyai karakteristik yang berbeda antara satu cluster dengan cluster lainnya yaitu pendapatan dari pajak daerah dan populasi dari setiap provinsi. Pemilihan jumlah cluster tersebut terbentuk berdasarkan nilai between SS / total SS tertinggi yaitu 99.1 persen ketika jumlah cluster sama dengan enam.
- Cluster 1 terdiri dari empat provinsi, yaitu Banten, Sumatera Utara, Kalimantan Timur, dan Jawa Tengah dimana digolongkan ke dalam daerah “Cukup Mandiri”. Penggolongan ini didapatkan karena jumlah populasi yang menengah diimbangi dengan pendapatan (pajak daerah) yang menengah juga.
- Cluster 2 terdiri dari satu provinsi, yaitu DKI Jakarta dimana digolongkan ke dalam daerah “Sangat Mandiri”. Penggolongan ini didapatkan karena jumlah populasi yang sedikit namun memiliki pendapatan (pajak daerah) yang sangat tinggi.
- Cluster 3 terdiri dari sembilan provinsi, yaitu D I Yogyakarta, Nusa Tenggara Barat, Sumatera Barat, Sulawesi Utara, Nanggroe Aceh Darussalam, Kepulauan Riau, Kalimantan Tengah, Kalimantan Barat, Jambi dimana digolongkan ke dalam daerah “Bergantung”. Penggolongan ini didapatkan karena jumlah populasi yang sedikit dan juga memiliki pendapatan (pajak daerah) yang kecil.
- Cluster 4 terdiri dari dua belas provinsi, yaitu Nusa Tenggara Timur, Sulawesi Tengah, Kepulauan Bangka Belitung, Sulawesi Barat, Maluku Utara, Kalimantan Utara, Papua, Sulawesi Tenggara, Maluku, Bengkulu, Papua Barat, Gorontalo dimana digolongkan ke dalam daerah “Sangat Bergantung”. Penggolongan ini didapatkan karena jumlah populasi yang sedikit ditambah juga dengan pendapatan (pajak daerah) yang sangat kecil.
- Cluster 5 terdiri dari enam provinsi, yaitu Bali, Sumatera Selatan, Lampung, Kalimantan Selatan, Sulawesi Selatan, Riau dimana digolongkan ke dalam daerah “Cukup Bergantung”. Penggolongan ini didapatkan karena jumlah populasi yang



menengah namun memiliki pendapatan (pajak daerah) yang terbelah menengah ke bawah.

- Cluster 6 terdiri dari dua provinsi, yaitu Jawa Timur dan Jawa Barat dimana digolongkan ke dalam daerah “Mandiri”. Penggolongan ini didapatkan karena jumlah populasi yang besar diimbangi dengan pendapatan (pajak daerah) yang juga tinggi.

Maka dari itu disarankan bahwa Pemerintah harus bisa memberikan prioritas dalam menyusun strategi pembangunan dimulai dari provinsi yang paling memerlukan bantuan yaitu pada cluster “Sangat Bergantung”. Pemerataan jumlah penduduk mungkin bisa dipertimbangkan untuk dapat memicu peningkatan kemandirian dari setiap pemerintah daerah. Hasil visualisasi menggunakan peta yang interaktif diharapkan dapat mempermudah pemerintah untuk mengatur strategi ke depannya.

Untuk penelitian selanjutnya diharapkan tidak hanya meneliti dengan variabel pajak daerah dan populasi, melainkan berdasarkan variabel lainnya yang termasuk dalam Pendapatan Asli Daerah seperti retribusi daerah, bagi hasil pajak, bagi hasil bukan pajak, dan lainnya. Diharapkan untuk ke depannya, objek penelitian tidak hanya pada Pemprov, melainkan pada seluruh Pemkot dan Pemkab di Indonesia.

Hasil dari project yang telah dibuat telah diupload pada link berikut :

<https://github.com/astridperdana/Pemrograman-Integratif>

## References