BAB 4

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Tahapan penelitian

4.1.1 Study Literatur

Tahapan penelitian ini membahas mengenai sumber ilmu yang digunakan terutama ilmu yang berkaitan dengan data, sistem rekomendasi dan teori yang sering digunakan.

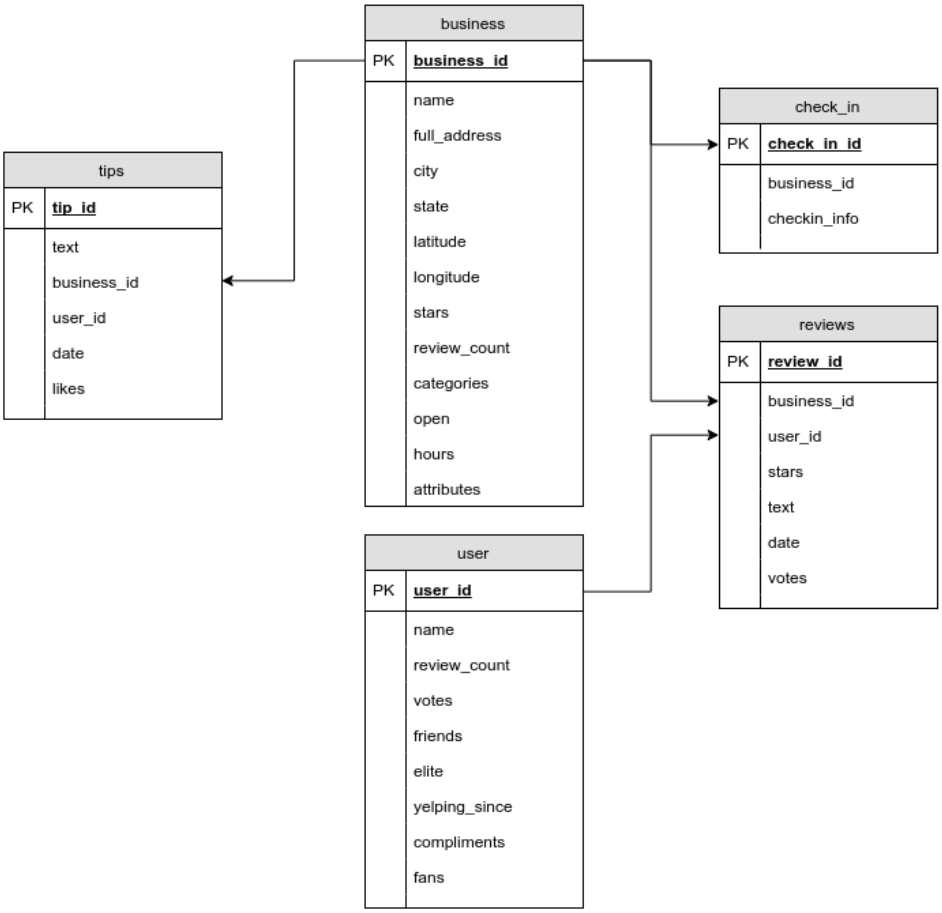
4.1.2 Pengumpulan Data

Pada tahapan penelitian ini, peneliti melakukan setup database dan melakukan import data. Adapun sumber data yang digunakan berasa dari yelp.com, sebuah social media yang didalamnya membahas *Business Review* (pembahasan soal tempat usaha / bisnis) yang bertemakan media sosial sehingga tiap-tiap pengguna dapat melakukan penilaian terhadap tempat usaha atau bisnis dan dapat dilihat oleh teman / orang-orang yang mengunjungi halaman website tersebut, adapun tahapan dalam pengumpulan data tersebut yaitu :

1) Proses download data : Data yang digunakan dalam keseluruhan penelitian dapat di download secara bebas dan legal pada sistus https://www.yelp.com/dataset_challenge.

2) File yang di download pada point 1 berbentuk json, untuk melakukan import ke database agar nanti dapat digunakan maka digunakan converter. Converter yang digunakan yaitu file berekstensi python (Lampiran 1 json_to_csv_converter.py).

3) Setelah struktur json pada lampiran 2 sampai dengan lampiran 6 dikonversi menggunakan converter dan file csv tersebut di import ke database maka didapat kumpulan tabel berelasi sebagai berikut,



Gambar 4.1 Relasi Antar Tabel

Pada Gambar 4.1 menunjukan relasi antar tabel, tabel tersebut dibagikan oleh yelp.com yang didalamnya terdapat informasi mengenai *local business* dari 10 kota yang ada di 4 negara, secara umum tabel diatas memiliki informasi seputar *local business* dan tanggapan dari *reviewer* yang sudah pernah singgah pada tempat tersebut.

Berikut merupakan keterangan untuk tabel-tabel yang ada pada gambar 4.1

1) Tabel User

Tabel user digunakan untuk menampung data *user*, *primary key* dari tabel ini adalah user_id. Struktur tabel yang lebih lengkap dapat dilihat pada Tabel 4.1

Tabel 4.1 Tabel User

Nama Field	Tipe Data	Kunci	Ukuran	Keterangan
user_id	varchar	PK	255	NOT NULL
Name	varchar		255	NOT NULL
review_count	integer			
average_stars	numeric			
Votes	json			
Friends	json			
Elite	json			
yelping_since	varchar			NOT NULL
Compliments	json			
Fans	integer			

2) Tabel *Business*

Tabel *business* digunakan untuk menampung data *local business*, *primary key* dari tabel ini adalah business_id. Struktur tabel yang lebih lengkap dapat dilihat pada Tabel 4.2

Tabel 4.2 Tabel Business

Nama Field	Tipe Data	Kunci	Ukuran	Keterangan
business_id	varchar	PK	255	NOT NULL
Name	varchar		255	NOT NULL
full_address	varchar			NOT NULL
city	varchar			NOT NULL
state	varchar			NOT NULL
latitude	numeric		(12, 7)	
longitude	numeric		(12, 7)	
stars	numeric		(12, 7)	
review_count	integer			
categories	json			
open	boolean			NOT NULL
hours	json			
attribute	json			

3) Tabel reviews

Tabel *reviews* digunakan untuk menampung data *review*, *primary key* dari tabel ini adalah review_id. Struktur tabel yang lebih lengkap dapat dilihat pada Tabel 4.3

Tabel 4.3 Tabel *Reviews*

Nama Field	Tipe Data	Kunci	Ukuran	Keterangan
review_id	varchar	PK	255	NOT NULL
business_id	varchar	FK	255	NOT NULL
user_id	varchar	FK		NOT NULL
stars	numeric			NOT NULL
text	text			
date	date			NOT NULL
votes	json			

4) Tabel tips

Tabel *tips* digunakan untuk menampung data *user* yang memberikan tip atau komentar baik ke *local business* yang dikunjunginya, *primary key* dari tabel ini adalah tip_id. Struktur tabel yang lebih lengkap dapat dilihat pada Tabel 4.4

Tabel 4.4 Tabel Tips

Nama Field	Tipe Data	Kunci	Ukuran	Keterangan
review_id	varchar	PK	255	NOT NULL
business_id	varchar	FK	255	NOT NULL
user_id	varchar	FK		NOT NULL
stars	numeric			NOT NULL
text	text			
date	date			NOT NULL

5) Tabel checkin

Tabel *checkin* digunakan untuk menampung *data check-in* atau waktu buka pada *local business, primary key* dari tabel ini adalah checkin_id. Struktur tabel yang lebih lengkap dapat dilihat pada Tabel 4.5

Tabel 4.5 Tabel *Check-in*

Nama Field	Tipe Data	Kunci	Ukuran	Keterangan
checkin_id	varchar	PK	255	NOT NULL
business_id	varchar	FK	255	NOT NULL
checkin_info	json			NOT NULL

4.2 Pembahasan

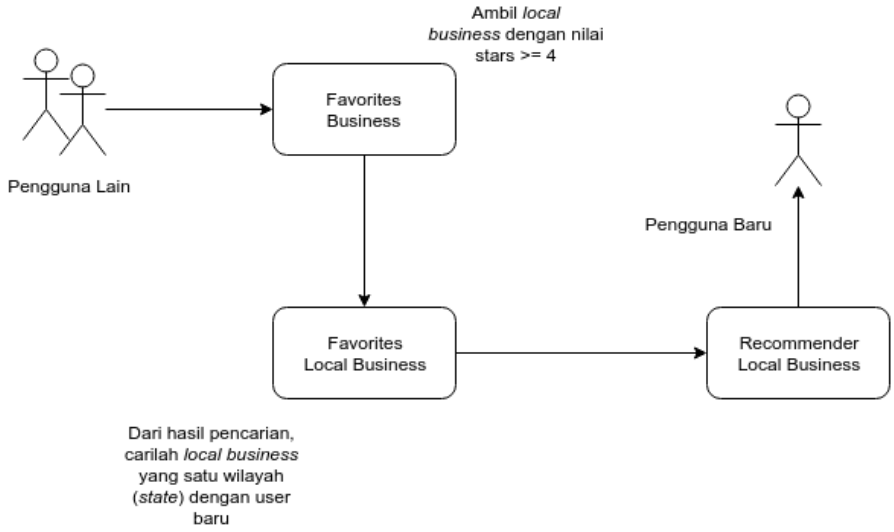
4.2.1 Proses Pembutan Daftar Rekomendasi – User Baru

Sampel data yang diambil yaitu user dengan tahun bergabung 2015 dan memiliki banyak review. Data sampel ini akan dilakukan pergeseran periode dan membuat seakan-akan data sampel ini belum terisi kecuali data user. Table 4.6 merupakan informasi dari data sampel yang akan digunakan :

Tabel 4.6 Data Sampel - *Data User* Baru

No	*Attribute*	*Value*
1.	user_id	"7dv_YsbEgxXj7AnrizdfBQ"
2.	name	"Lauren"
3,	review_count	318
4.	yelping_since	2015-06
5.	user_state	NV
6.	average_stars	4.22

Sistem rekomendasi yang terbentuk dari *demographic* dan *ratings* mengambil ratings atau stars dengan nilai >= 4 yang artinya tempat tersebut memiliki rating baik serta diurutkan berdasarkan jumlah *reviewer* terbanyak.



Gambar 4.2 Demographic and Ratings algorithm
for recommender system

Pada Gambar 4.2 dijelaskan Bahwa data yang diambil untuk sistem rekomendasi berasal Local business yang memiliki state / daerah yang terdekat dengan user baru, serta memiliki penilaian baik dari user lain yang tedapat dalam sistem.

Proses *demographic* dan keperluan untuk pengujian menggunakan MAE, berikut *query* yang dieksekusi dengan tujuan mengambil nama *local business, average reviews*, *user reviews* dan selisi antara *average reviews* dan *user reviews.*

```
SELECT bus.name, bus.stars, rev.stars as user_stars, bus.stars -
rev.stars as selisih
  FROM business bus
  JOIN reviews rev ON (rev.business_id = bus.business_id)
 WHERE rev.user_id = '7dv_YsbEgxXj7AnrizdfBQ'
ORDER BY random()
LIMIT n;
```

Persamaan 4.1 Query Demographic

Pada persamaan 4.1 diambil user dengan ID = 7dv_YsbEgxXj7AnrizdfBQ yang merupakan data sampel yang ditentukan sebelumnya. Sedangkan Limit n menunjukan jumlah data yang akan di uji pada tiap MAE yaitu 20, 50, dan 100 data. Perbandingan jumlah data nantinya akan dihitung dan diamati, apakah terjadi penurunan pada nilai MAE atau tidak, hal tersebut menentukan kualitas dari metode yang digunakan saat pembuatan sistem rekomendasi.

Tabel 4.7 Hasil dari sistem rekomendasi – User Baru

No	***Name***	***City***	***State***	***Stars***	***Reviewer***
1.	Mon Ami Gabi	Las Vegas	NV	4	5642
2.	Earl of Sandwich	Las Vegas	NV	4.5	4452
3.	Bacchanal Buffet	Las Vegas	NV	4	4390
4.	Gordon Ramsay BurGR	Las Vegas	NV	4	3811
5.	Hash House A Go Go	Las Vegas	NV	4	3014
6.	The Cosmopolitan of Las Vegas	Las Vegas	NV	4	2988
7.	Bouchon at The Venetian	Las Vegas	NV	4	2922
8.	Lotus of Siam	Las Vegas	NV	4	2818
9.	Secret Pizza	Las Vegas	NV	4	2758
10.	Mesa Grill	Las Vegas	NV	4	2442

Tabel 4.7 merupakan hasil dari sistem rekomendasi yang di tujukan untuk user baru dengan pendekatan *demographic* didalamnya terdapat nama *local business*, kota, provinsi, nilai dan banyaknya *reviewer.* Ke – 10 nama *local business* tersebut merupakan hasil generate dari pendekatan *demographic* yang nantinya pada tiap iterasi akan berbeda-beda.

Tabel 4.8 Data Sampel – Real User Review, Data User Baru

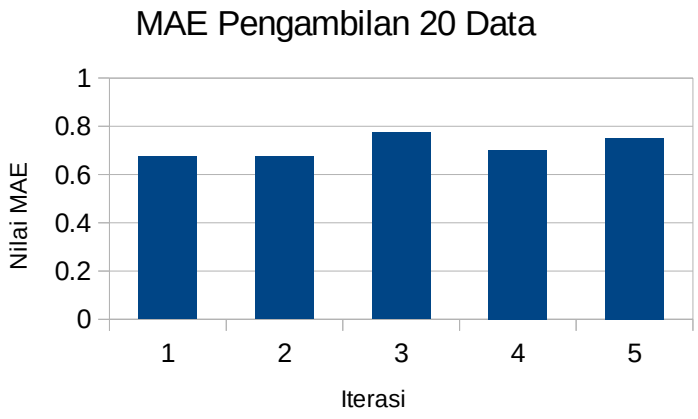
No	***Business name***	***State***	***City***	***Stars***	***Review Date***
1.	Todd's Unique Dining	NV	Las Vegas	5	2015-06-07
2.	I Love Sushi	NV	Henderson	5	2015-06-07
3.	Ferraro's Italian Restaurant & Wine Bar	NV	Las Vegas	5	2015-06-20
4.	Paymon's Mediterranean Cafe & Lounge	NV	Las Vegas	5	2015-08-02
5.	Emeril's New Orleans Fish House	NV	Las Vegas	5	2015-11-30
6.	Penn & Teller	NV	Las Vegas	5	2015-07-05
7.	**Mesa Grill**	**NV**	**Las Vegas**	**5**	**2015-12-06**
8.	Palms Casino Resort	NV	Las Vegas	5	2015-07-05
9.	**Mon Ami Gabi**	**NV**	**Las Vegas**	**5**	**2015-06-20**
10.	Tom Colicchio's Craftsteak	NV	Las Vegas	5	2015-06-22
11.	Chevys Fresh Mex	NV	Henderson	3	2015-10-12
12.	Le Macaron	NV	Henderson	2	2015-11-21
13.	Ventano Italian Grill & Seafood	NV	Henderson	2	2015-06-07
14.	Stratosphere Buffet	NV	Las Vegas	2	2015-12-06
15.	Il Fornaio	NV	Las Vegas	2	2015-06-28
16.	Bacio	NV	Las Vegas	2	2015-06-22
17.	Sammy's Restaurant & Bar	NV	Henderson	2	2015-07-05
18.	Hank's Fine Steaks	NV	Henderson	1	2015-12-06
19.	GoWireless Verizon Premium Retailer	NV	Henderson	1	2015-12-03
20.	Cactus Garage Doors	NV	North Las Vegas	1	2015-07-18

Note : Baris yang bergaris tebal merupakan kecocokan dari hasil rekomendasi dengan aktifitas user.

4.2.2 Penerapan metode dan pengujian *demographic*

Untuk mengetahui seberapa akurat pendekatan yang digunakan, maka dapat di uji dengan menggunakan persamaan MAE sebagai berikut :

1. Asumsi user sampel belum memiliki data review karena dianggap user baru.
2. Data training mulai dari 20, 50, dan 100 data diambil secara acak.
3. Tiap pengambilan data diambil 5 iterasi.

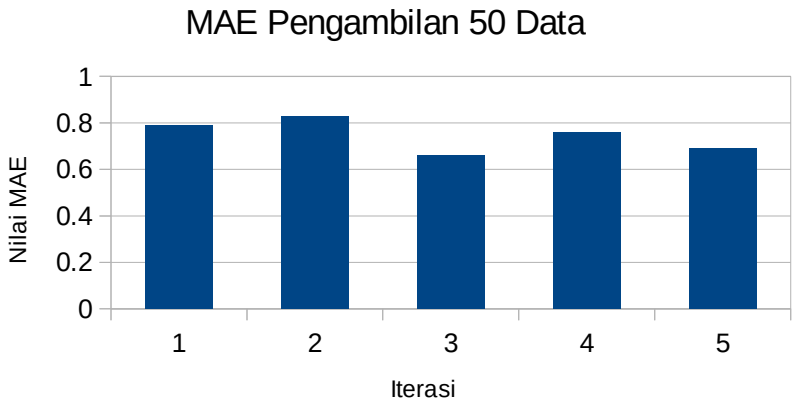


Gambar 4.3 Demographic Filtering MAE 20 Data

Gambar 4.3 Menujukan grafik MAE yang diperoleh dari pengambilan 20 Data secara acak. Tabel 4.9 merupakan detail dan rata-rata nilai untuk MAE pada pengambilan 50 data *recommender* secara acak.

Tabel 4.9 Hasil MAE 20 Data

20 Data Training		
Iterasi	Nilai MAE	Avg MAE
1	0.675	0.715
2	0.675	
3	0.775	
4	0.7	
5	0.75	



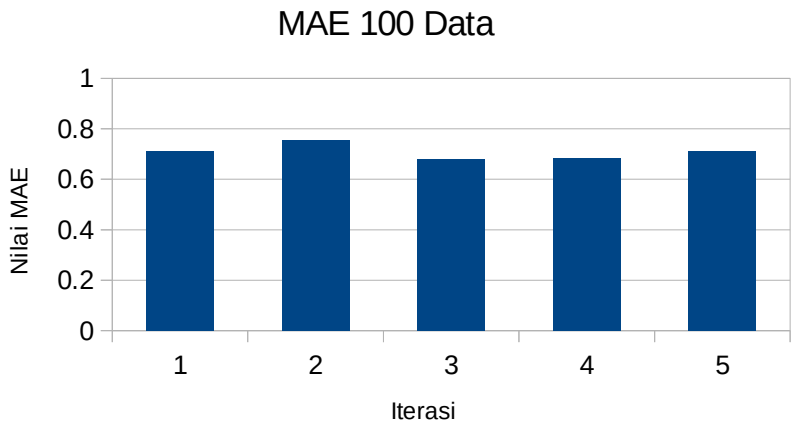
Gambar 4.4 Demographic Filtering MAE Pengambilan 50 Row

Gambar 4.4 menunjukan hasil grafik nilai MAE dari pengambilan 50 data

recommender secara acak. Tabel 4.10 merupakan detail dan rata-rata nilai untuk MAE pada pengambilan 50 data *recommender* secara acak.

Tabel 4.10 Hasil MAE 50 Data

50 Data Training		
Iterasi	Nilai MAE	Avg MAE
1	0.79	0.746
2	0.83	
3	0.66	
4	0.76	
5	0.69	



Gambar 4.5 Demographic Filtering MAE 100 Data

Gambar 4.5 menunjukan hasil grafik nilai MAE dari pengambilan 100 data *recommender* secara acak. Tabel 4.11 merupakan detail dan rata-rata nilai untuk MAE pada pengambilan 100 data *recommender* secara acak.

Tabel 4.11 Hasil MAE 100 Data

100 Data Training		
Iterasi	Nilai MAE	Avg MAE
1	0.71	0.708
2	0.755	
3	0.68	
4	0.685	
5	0.71	

Dari data MAE menujukan bahwa hasil dari demograhic filtering terjadi penurunan nilai MAE ketika data training yang diambil sebanyak 100 data. Hal tersebut menunjukan bahwa makin banyak data yang dibandingkan maka nilai MAE semakin kecil artinya nilai kesalahan semakin kecil, sehingga dapat menghasilkan sistem rekomendasi yang baik meskipun data yang dimiliki untuk user baru cenderung tidak lengkap atau sedikit.

4.2.3 Hasil Akhir Sistem Rekomendasi User Baru

Dengan mengacu pada MAE yang sama dengan atau mendekati 0 dan memiliki *rating* >= 4 maka didapat hasil *recommender system* dari data sampel yaitu sebagai berikut :

Tabel 4.12: Hasil *Recommender System* Pendekatan *Demographic*

Local Business Name	***Average ratings***
YEE: Yelp's BLT SnackDown	5
Totally Olive	5
Pizza Rock	4
Tom's Urban	4
Gimme Some Sugar	4
Bernard's Bistro	4
The Volcano	4
Samurai Japanese Steakhouse & Sushi	4
Flame Kabob	4
Lucky Foo's Restaurant & Bar	4

4.2.4 Proses Pembutan Data Rekomendasi – User Dengan Transaksi

Tabel 4.13 Data Sampel - User Dengan Transaksi

No	Attribute	Value
1	user_id	"1p3d__fuRRCDXfbS1Tq0wA"
2	Name	"Shila"
3	review_count	8529
4	average_stars	3.74

Dari data sampel (Tabel 4.13) tersebut diambil transaksi yang memiliki jumlah rating tertinggi yang di buat oleh *user*, Berdsarakan Data yang ada user mengisi review dengan *rating* yang tinggi untuk *data business* sebagai berikut :

Tabel 4.14 Hasil *data review* dari *data sampel*

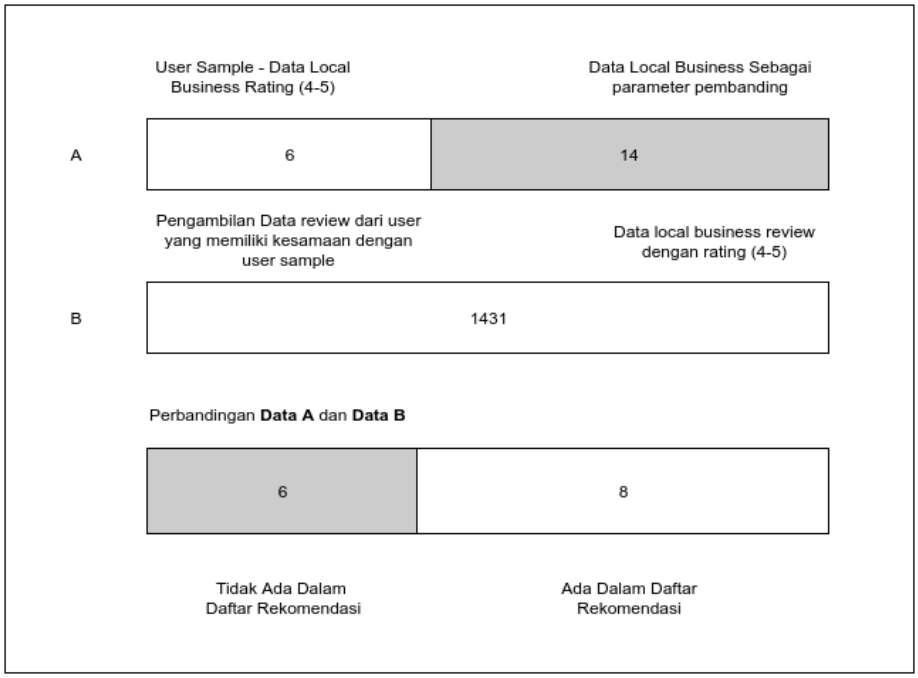
No	Name	State	Stars	Date
1.	M & M's World	NV	5	2013-04-26
2.	House of Blues	NV	5	2013-04-26
3.	MGM Grand Buffet	NV	4	2013-04-26
4.	Everything Coca-Cola	NV	4	2013-04-26
5.	FAO Schwarz	NV	5	2013-04-26
6.	Caesars Palace Las Vegas Hotel & Casino	NV	4	2013-04-26

Dari sebagian data diatas, kemudian didapat beberapa user yang memiliki kesamaan dengan Data sampel User :

Tabel 4.15 Daftar User yang Berkaitan Dengan Data sampel

No	Name	Yelping_since	average_stars
1.	Andrew	2005-07	3.69
2.	Nijole	2011-11	3.73
3.	Jennifer	2009-05	3.96
4.	Jando	2006-11	3.33
5.	Stefany	2011-06	3.27
6.	Anthony	2009-03	3.23
7.	Stephy	2006-05	3.88
8.	Steph	2008-07	4.13
9.	Jana	2009-08	3.91
10	Sunil	2009-01	3.44

Dari 10 *data user* diatas kemudian didapat hasil *local business* mana yang memiliki rating baik (4-5) yang di review oleh ke 10 data user tersebut yang tentunya selain dari ke 6 data *local business* yang di *review* oleh data sampel user.



Gambar 4.6 Hasil Dari Collaborative Filtering

Dari Gambar 4.6 dapat disimpulkan bahwa persentase keberhasilan pembuatan sistem rekomendasi dengan *collaborative-filtering* sebesar 58% (8 dari 14) faktor penentu besarnya nilai keberhasilan suatu daftar rekomendasi dapat dilihat dari beberapa parameter pembangungnya.

Parameter tersebut seperti data sampel yang diambil, rentang waktu pengambilan data untuk data sampel, dan banyaknya data dalam semesta sistem tersebut.

Berikut tahapan dalam mendapatkan rekomendasi menggunakan *collaborative-filtering* :

1) Tampilkan *local business* dengan nilai rate tinggi (>= 4) yang di miliki oleh data sampel :

```
SELECT b.business_id, b.name, b.state, r.stars, r.date FROM reviews
r
JOIN business b on (b.business_id = r.business_id)
 WHERE user_id = '1p3d__fuRRCDXfbS1Tq0wA' AND
 r.stars >= 4
ORDER BY r.date asc;
```

Persamaan 4.2 *Query collaborative-filtering* (1)

2) Ambil 20 % dari jumlah data yang tampil pada persamaan 4.2

3) Dari data 20% pada point 2, cari beberapa *user* dengan kemiripan yang serupa, yaitu user yang memiliki ketertarikan pada *local business* pada point 2 selain data sampel.

```
SELECT DISTINCT r.user_id
  FROM reviews r
  JOIN business b on (b.business_id = r.business_id)
 WHERE b.business_id IN (
  'ZgWUM594Yqk52IvBgoIrdA', 'M2SjLXsuAy5RNKwxaA5E1g',
'c12WDNrP2wyQgbnEYDdxCw',
  'jcI4ZPF3-hSUPsnoqRHO4w', 'aBEg8BGc_4ZerEfzBeVHZQ',
'z3SyT8blMIhsZNvKJgKcRA'
) AND r.stars >= 4;
```

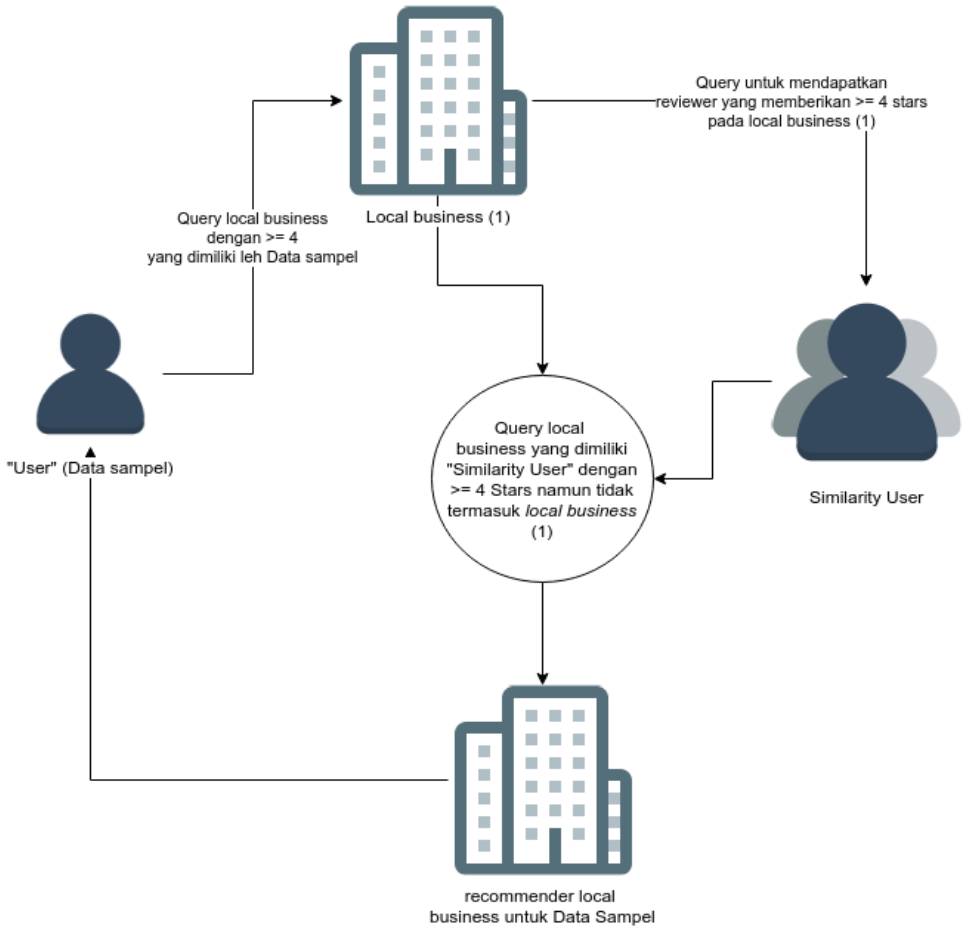
Persamaan 4.3 *Query collaborative-filtering* (2)

4) Pada tahap berikutnya, cari *local business* selain poin 2 dengan bantuan data user pada poin 3 yang tentunya memiliki *stars* atau *rate* >= 4. Tahapan ini merupakan tahapan akhir dan data pada poin ini dapat direkomendasikan pada *user* yang sudah memiliki transaksi.

```
SELECT DISTINCT b.business_id, b.name, b.state, r.stars, r.date
  FROM reviews r
  JOIN business b on (b.business_id = r.business_id)
 WHERE r.user_id IN (
      '7FuLnS_-b79GG-33mwLaMg', 'sKtmrRvPMn5GRTnHYd41CA',
'pEVf8GRshP9HUkSpizc9LA',
      'vyfsQo-estP8EfiIFMsL6g', '9A2-wSoBUxlMd3LwmlGrrQ',
'xNb8pFe99ENj8BeMsCBPcQ',
      'qbfQRHLvZk5WSkKY0l_lMw', 'dQhd0Wfn59TYWf2GVukEww',
'7FRi5k2A8fXECbBFwno1KA',
      'KfC3kgw-EzI3wECydlNYLA'
) AND b.business_id NOT IN (
  'ZgWUM594Yqk52IvBgoIrdA', 'M2SjLXsuAy5RNKwxaA5E1g',
'c12WDNrP2wyQgbnEYDdxCw',
  'jcI4ZPF3-hSUPsnoqRHO4w', 'aBEg8BGc_4ZerEfzBeVHZQ',
'z3SyT8blMIhsZNvKJgKcRA'
) AND r.stars >= 4;
```

Persamaan 4.4 *Query Collaborative-Filtering (3)*

Untuk lebih jelasnya proses atau tahapan-tahapaan pada *collaborative-filtering* dapat di lihat pada gambar 4.7

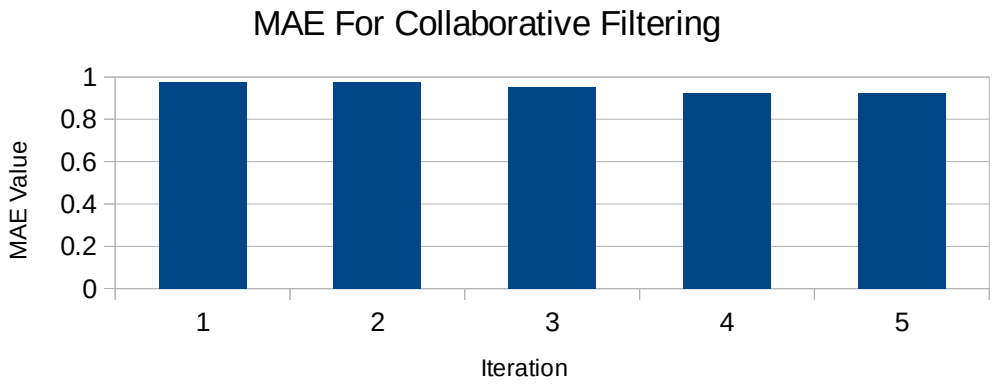


Gambar 4.7 *Collaborative Filtering Process*

4.2.5 Pengujian Hasil *Collaborative Filtering*

Pada tahapan ini diuji sebarapa besar nilai kemiripan yang dimiliki antara satu *local business* dengan *local business* yang lain :

1. Ambil data sampel yaitu *user* yang sudah memiliki beberapa transaksi pada sistem (*review*) ambil 20 % untuk *data testing* dan 80% untuk *data training*.
2. Ambil local business yang di *review* oleh data sampel dan lakukan pencarian terhadap beberapa *user* lain yang memiliki ketertarikan yang sama dengan data sampel.
3. Dari beberapa user tersebut, ambil *local business* yang di beri *rating* baik oleh user tersebut.
4. Bandingkan *local business* yang di rekomendasikan pada poin 3 dengan *local business* yang di beri *rating* baik oleh data sampel.



Gambar 4.8 MEA Collaborative Filtering

Gambar 4.8 merupakan hasil MAE dari pendekatan *collaborative-filtering*, dan tabel 4.16 merupakan detail dan nila rata-rata dari MAE dengan pendekatan *collaborative-filtering*

Tabel 4.16 *MAE Collaborative Filtering*

Data Training		
Iterasi	Nilai MAE	Avg MAE
1	0.975	0.95
2	0.975	
3	0.95	
4	0.925	
5	0.925	

4.2.6 Hasil Rekomendasi Menggunakan *Adjust Cosine* dan *Weighted Sum*

Berikut meruapakan tahapan mendapatkan hasil dari sistem rekomendasi dengan pendekatan *collaborative-filtering* dengan menggunakan *Adjust cosine* dan *weighted sum* :

1. Ambil sampel data berupa dua local business yang didapatkan dari *user data* sampel pada Tabel 4.13.
2. Cari beberapa user dengan kemiripan berdasarkan *rating* pada *local business* tersebut, dalam kasus ini empat user yang diambil.
3. Cari lima local business lain secara acak beserta rating yang dimiliki oleh user pada point 2 dan *user data* sampel.
4. Uji dengan persamaan *Adjust cosine* dan *Weighted sum.*

Tabel 4.17 Skenario Rating

	House of Blues	Popped	Fashion Show	M & M's World	Ev. Coca-Cola	The Forum Shops	*Avg Ratings*
Dana	4			3	5		4
Scott	4		4	4	4	4	4
David	4	5	3	3	4	4	3.83
Bethany	5	4	4	3		5	4.2
Shila	5			5			5

Tabel 4.17 merupakan skenario rating untuk mendapatkan hasil sistem rekomendasi yang akan ditampilkan untuk data sampel (“Shila”). Tahap selanjutnya yaitu menghitung *similarity* antar *local business* dengan persamaan *adjust cosine.*

$$S(i,j) = \frac{\sum_{u \in U} (Ru,i - \hat{R}i)(Ru,j - \hat{R}j)}{\sqrt{\sum_{u \in U} (Ru,i - \hat{R}i)^2} \sqrt{\sum_{u \in U} (Ru,j - \hat{R}j)^2}}$$

$$S(i,j) = \frac{\sum_{u \in U} (Ru,i - \hat{R}u)(Ru,j - \hat{R}u)}{\sqrt{\sum_{u \in U} (Ru,i - \hat{R}u)^2} \sqrt{\sum_{u \in U} (Ru,j - \hat{R}u)^2}}$$

Tabel 4.18 Hasil Kesamaan Antar *Local Business*

Local Bus 1	***Local Bus 2***	**Nilai**
House of Blues	Popped	0.19
House of Blues	Fashion Show	-0.43
House of Blues	Ev Coca Cola	0.17
House of Blues	The Forum Shops	1.00
M & M's World	Popped	-0.42
M & M's World	Fashion Show	0.75
M & M's World	Ev Coca Cola	-0.87
M & M's World	The Forum Shops	-0.92

Tabel 4.18 menujukan bahwa hasil dari *adjust cosine* ada diantara -1 hingga 1 jika nilai mendekati 1 maka kedua *local business* tersebut memiliki kemiripan yang kuat. Sebaliknya jika mendekati -1 maka *local business* tersebut saling berlawanan. Kesimpulannya sistem rekomendasi hanya menampilkan nilai yang > 0

Setelah memperoleh nilai kesamaan antar *local business*, nilai yang lebih besar dari 0 akan digunakan untuk mendapatkan nilai prediksi, karena nilai tersebut dianggap sebagai batas bawah keterhubungan antar *local business*. Untuk memperoleh nilai prediksi dari suatu *local business* yang belum di rating oleh *users*, digunakan persamaan *weighted sum*.

$$P(a, j) = \frac{\sum_{i \epsilon I} (Ra,i \cdot Si,j)}{\sum_{i \epsilon I} |Si,j|}$$

Tabel 4.19 Hasil Prediksi *Collaborative Filtering*

Local Business	***Predicting value***
Popped	-2.49
Fashion Show	1.03
Ev Coca Cola	-3.37
The Forum Shops	0.21

Dari tabel 4.19 di dapat kesimpulan bahwa dari 4 local business yang menjadi sampel, hanya *Fashion show* dan The forum shops yang dapat direkomendasikan untuk user sampel (Shila).