

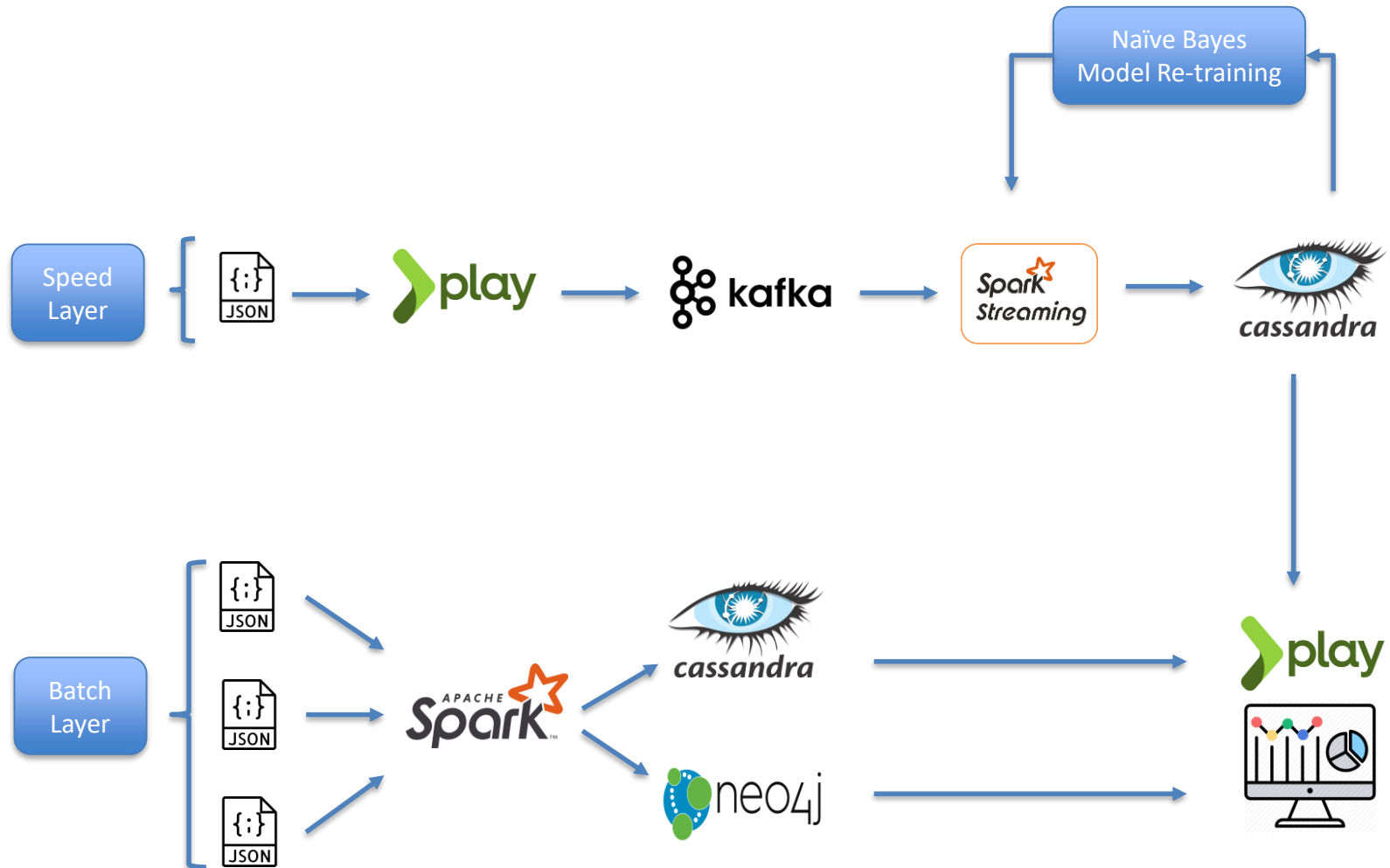
Sentiment Analysis on Yelp Dataset using Spark Structured Streaming and Kafka

- Team 7 : Akshay Patel, Keshav Vyas

Goals

- Processing Data at Scale, Can you imagine manually sorting through millions customer reviews, ratings, suggestions(tips)? There's just too much data to process manually. Sentiment analysis helps user to consolidate right data for a business.
- User can also find trending businesses which would help in finding the right business and potentially most popular and positive experience.
- Big data pipeline which continuously ingest source data and provide new insights for easier decision making for the user.

Flow Diagram



Use Case

(1) Click Submit
to Start Kafka



CSYE 7200
Sentiment Analysis of Yelp Dataset

Developed By Akshay Patel(1446491), Keshav Vyas(1409174)

Click on Submit button to start application

It Will Read Review Json File And Send Every Json Object As Kafka Message

Submit

Total Number of Review records processed: 2030798

Business Sentiment

User Id:

CxD0IDnH8gp9KXzpBHJYXw

Business Id:

5iSmZ00SrKU6EoXK_1M8Kw

Submit

(2) Click on
Submit for Use
Case 1 & 3

Trending Business

Business Category:

'Restaurants'

Location:

'Las Vegas'

Submit

(3) Click on
Submit for Use
Case 2



CSYE 7200


Sentiment Analysis of Yelp Dataset

Developed By **Akshay Patel(1446491)**, **Keshav Vyas(1409174)**


Business Name: 5iSmZO0SrKU6EoXK_1M8Kw

Business Score: 0.37254901960784315

Top Positive Review


 **UserName**

Definitely the best bang for the buck, easy location to major strip sights, conveniently to malls and foodcourt. Interesting pools, spacious casinos, friendly staff abound! Whether you've got kids or seniors or a group on a guy/gals getaway, this is the one resort in Vegas where I really felt all were welcomed! I've stayed at Wynn, mandalay, venetian, mgm, Flamingo, bellagio and this is my fav mid price resort. Go to the wynn if you want to splurge but be prepared for the inconvenience of WALKING/WAITING to get anywhere. Pros: Love the variety of pools, lazy river has current and wave pool makes a dip quite fun! Elevators are SO close to everything, if you got kids, are senior or just plain lazy like me, food, pool, casino and exits are all within 50 feet from the elevator. Wish more resorts were laid out with this convenience in mind. Rooms were standard for the \$100 price range, what do you expect? Cons: The resort fee has got to go! Marriott tried it and had to end it!

 **UserName**


I meannnnn 4 stars for Monte Carlo? If you've ever heard Karey R. and I talk about Vegas, as it is

Top Negative Review

 **UserName**

Can't wait for my last day here, which is tomorrow! This has all been a very stressful experience. I definitely suggest you rethink your visit here during their reconstruction. There is no front entrance to the check-in counter. We had to park so far and walk through several hallways just to get to the check in counter. At the check in counter was the shortest line with the longest wait, frustrating right? We finally got to our room and realized the toilet wasn't working right. The room isn't what you would expect. I used to think Monte Carlo was one of the nicer properties, today I know better. No bottled water as promised, no coffee maker, no mini fridge. I spent \$10 this morning just on some liquids to keep myself hydrated. I woke up this morning like a fish out of water, the a/c blew warm air all night. Talk about waking up like a raisin. I wish it was because I drank and partied so much last night, but no-it's because the a/c was bad. Oh, no pool of their own for me to jump in when it is 103 degrees out. To make matters worse, I woke up to the sounds of jack hammers and drilling. Not a very soothing way to wake up. I understand that it is not the hotels fault, they are under construction. I do think they should at least give their guests \$50 vouchers so we can at least go get some water and coffees and your property for he

Top Friends Reviews

 **UserName**

The hotel itself is clean but it's kinda out of date. The rooms had old wallpaper that's peeling off. There were 2 screws on the floor in our room and I was walking around barefoot. I came close to stepping on it. Luckily I saw it before I did. The cleaning crew must have gotten lazy.

 **UserName**

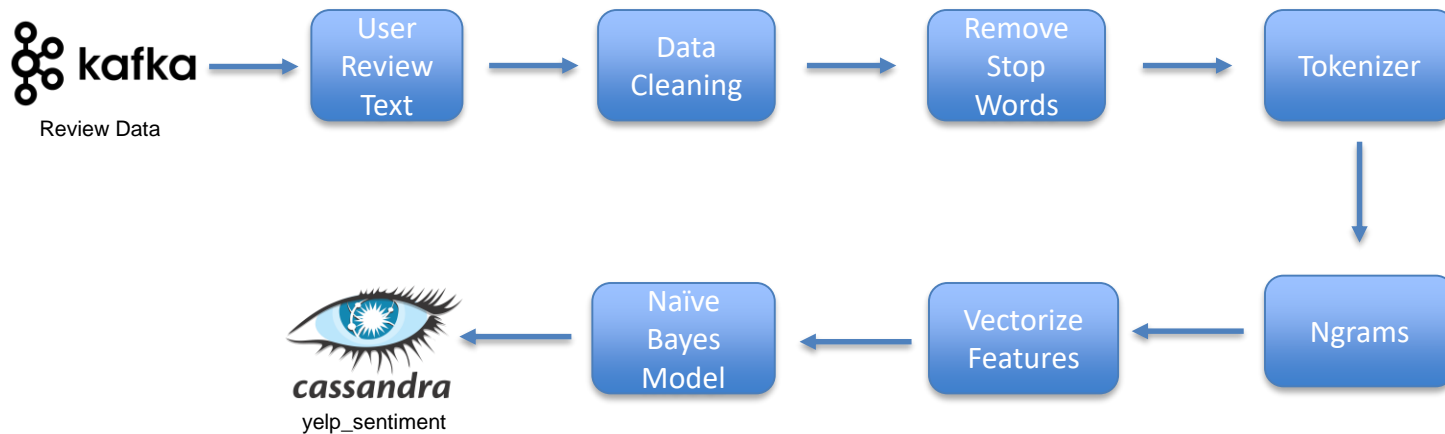
Hey, I remember these fixtures and these pastel paintings -- my mom decorated our home this way in 1992! Hello not-yet-retro, bleck! Everything started off predictably well - the check in service was quick, and the staff was friendly. But all went downhill when the light of day hit my room. No WiFi, and the wired in outlet was on the fritz. The room was not properly cleaned before my arrival. Sure, it was surface swept, but not well at all - the remains of a shopping spree's plastic price tags littered the floor between the beds, the toilet seat was wet, and there were strange black smears on two of the pillows and the outer comforter. Partner that up with a suspicious smell to the water, a shower fixture that could decapitate any royal and soaps that left me so dry I immediately peeled two layers of skin in an

'Restaurants'
'Las Vegas'

Trending Business

 Aloha Island Grill 'Las Vegas'	Business Score: 1.0
 The Corndog Company LV 'Las Vegas'	Business Score: 1.0
 Kosher On the Grill 'Las Vegas'	Business Score: 1.0
 Tsunami Asian Grill & Sushi Bar 'Las Vegas'	Business Score: 1.0
 Casanova Bites 'Las Vegas'	Business Score: 1.0
 Taqueria Santander 'Las Vegas'	Business Score: 1.0
 NoMad Bar 'Las Vegas'	Business Score: 1.0
 Redstone Grill 'Las Vegas'	Business Score: 1.0
 Chinatown Cuisine 'Las Vegas'	Business Score: 1.0
 Los Compadres 'Las Vegas'	Business Score: 1.0

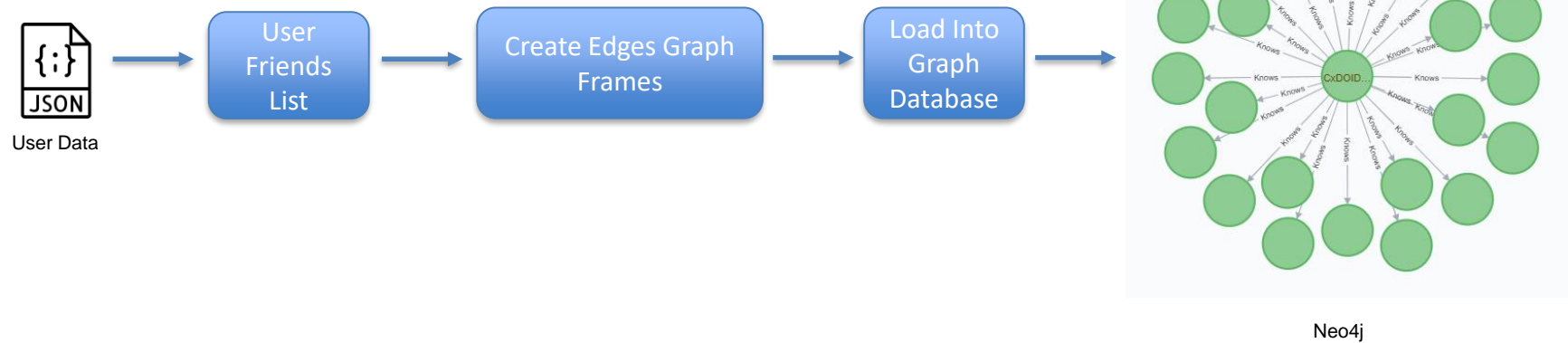
Review Sentiment Classification



Model Re-training



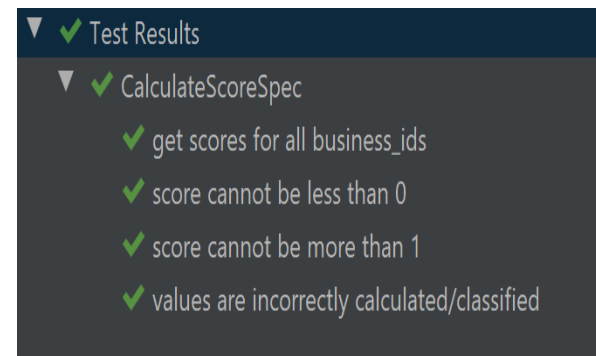
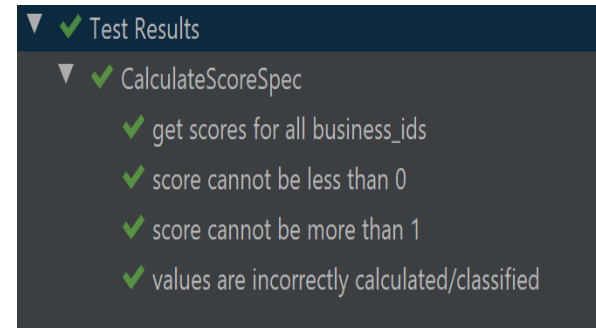
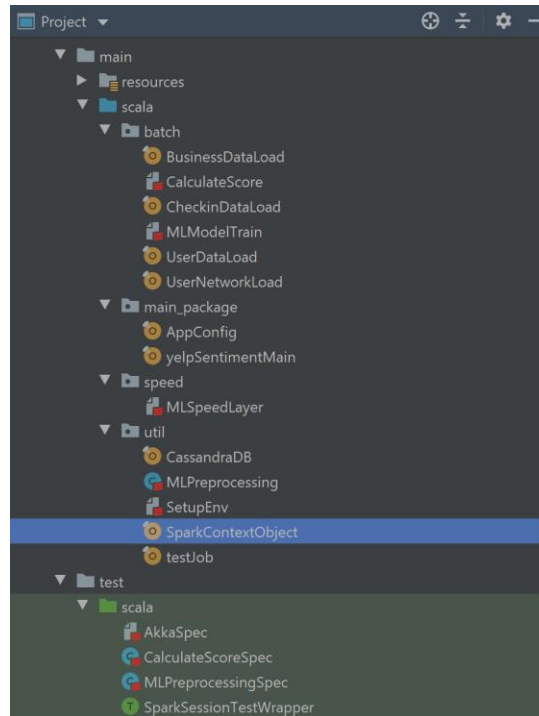
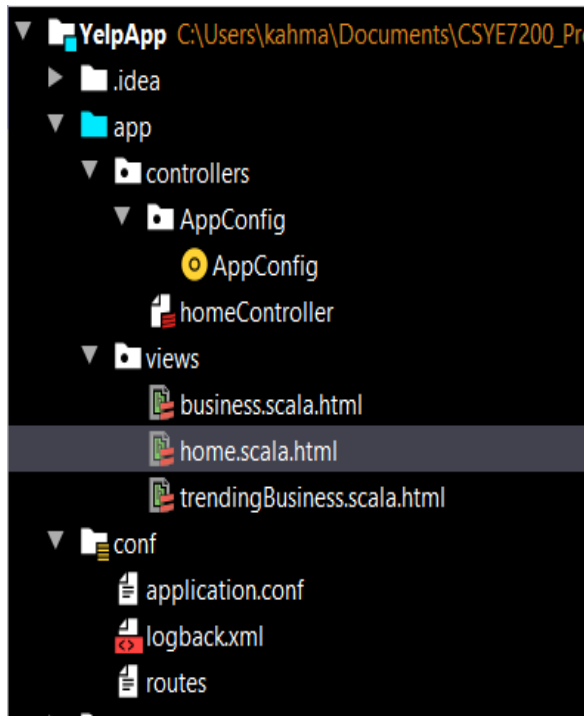
Create Friends' Network



Calculate Business Score



Programming in Scala



Coding Repository:

https://github.com/vyaskeshav/CSYE7200_Project

Acceptance Criteria

- Review transaction should be sent through Kafka and processed by spark streaming API in less than 5 seconds

Total Records: 2030798

Total Time: 827 Sec

Time per Record: 0.4 milliseconds

```
Elapsed time: 1sec
1
Elapsed time: 3sec
2
Elapsed time: 31sec
3
Elapsed time: 430sec
4
Elapsed time: 362sec
|
```

- Positive/Negative review classification with at least 70% accuracy

```
NB Test set accuracy = 0.753003003003003
```

- Process user's friends by a specific business and develop a user network in 5 seconds

Total Records: 563 friends

```
Elapsed time: 5sec
|
```

Thank
You