UCLA, MASDS, STATS-418-S25

😍How do you feel about the UCLA?

:Real Time Sentiment Analysis via Reddit data ✅

Hochan Son

Architectural & Key design pattern

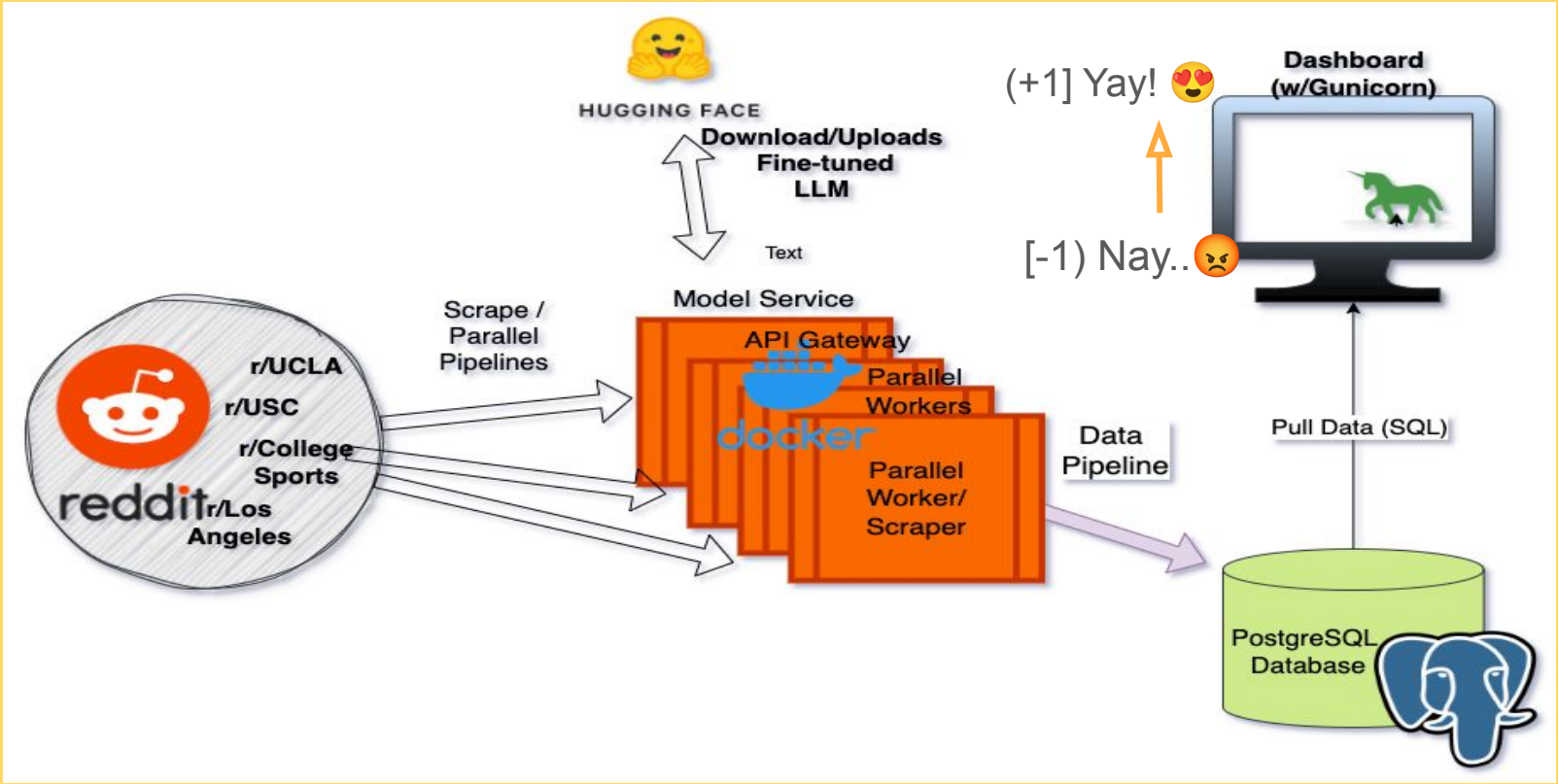
- **Circuit Breaker Pattern**:
  Automatically switches to **VADER** fallback when ML (LLM) models fail

- **Hot-swappable Models**:
  "Can change ML models without any system downtime"

- **Async Processing**:
  "5-10x performance improvement through non-blocking operations"

- **Container Orchestration**:
  "Full Docker deployment with one-command setup"

Data Pipeline Diagram



Pre-Objective

- What sentiments?

{ Hate, Frustrated, Super or emojis}

- Scrape data from Reddit / Subreddit channels

- Rate Limits (only 1000 post)

- Scoring system

- Keyword to any social sentiment

(i.e. USC, LA Olympic 2028, or Tariff)

- Building Real time Sentiment Analysis system

Observation

Name	VAEDAR	LLM
BaseModel	VAE (Variational autoencoder)	Distil-BERT
Algorithm	Encoder =[{hidden}]= Decoder	Transformer
Model Size	Small (500MB)	Large (1GB - 400GB)
Performance	60-70% 10-50 ms	97-99% 100+ ms

```
intensifiers = { 'absolutely': 0.293,
'completely': 0.293, 'extremely': 0.293,
'very': 0.293, 'quite': 0.2, 'rather': 0.2,
'somewhat': 0.2, 'slightly': 0.1 }

lexicon = {'amazing': 2.5, 'love': 3.2,
'hate': -2.7, 'terrible': -3.1, 'good': 1.9,
'bad': -2.5, 'awesome': 3.1,  'awful': -2.8}

emoticon_lexicon = { ':)': 2.1, ':-)': 2.1, ': )':
2.1, ':D': 2.6, ':-D': 2.6, ':(': -2.1, ':-(': -2.1, ':
(': -2.1, ':P': 1.4, ':-P': 1.4, ':o': -0.7, ':-o':
-0.7, '<3': 2.9, # heart '</3': -2.9, # broken
heart # ... many more emoticons }
```

$$\text{compound} = \frac{\text{sum(sentiment\_scores)}}{\sqrt{(\text{sum(sentiment\_scores})^2 + 15)}}$$

Implication / Future Improvement

Implications / Shortcoming

-Complex Design :

(Microservices, proposed 4 API services , Model-service, Dashboard-Service, Worker/Scarper, and API Gateway)

- Time allocation on Debugging up to 40%

-Better suit for bigger size team and ASYNC distributed System

Future Improvement

- CI/CD Build automation

- Automatic Testing feature

- Model Tine-tuning Learning rate Monitoring (Weight and Bias , WANDB)

Q & A

STAT-418

Thank you!!!!