Text Mining Applications

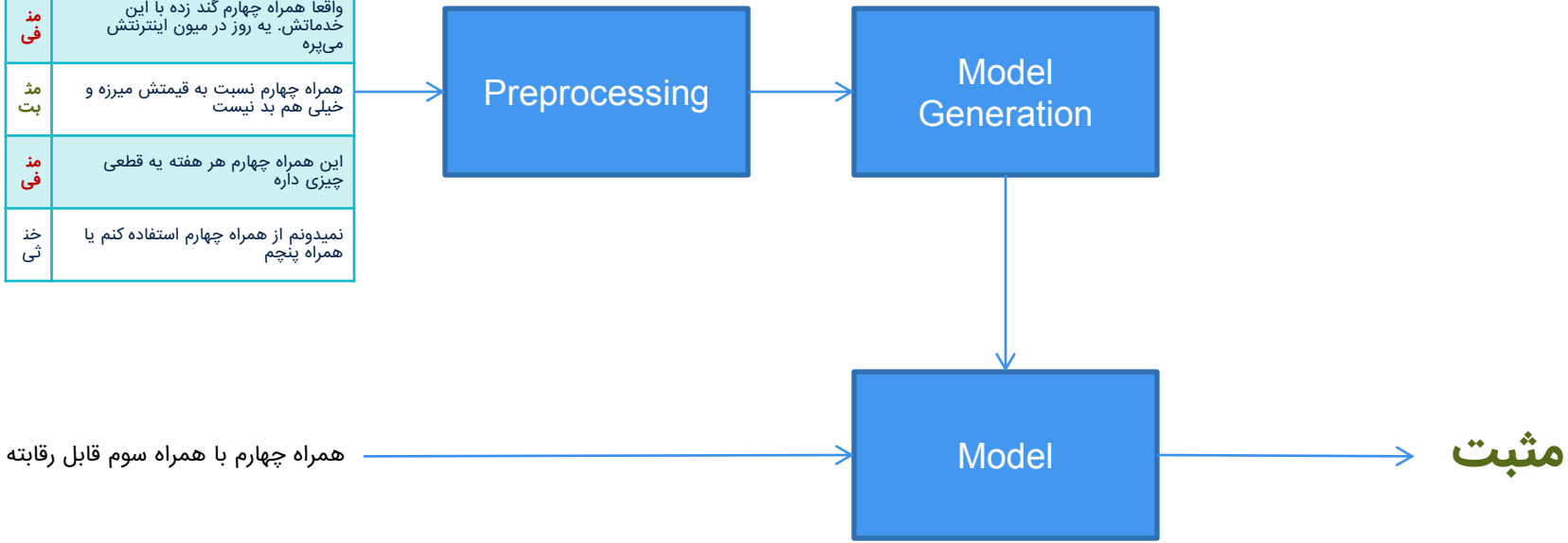
KhabarFarsi - ParsPack

What is Text Mining?

توییت	احساس
واقعا همراه چهارم گند زده با این خدماتش. یه روز در میون اینترنتش می‌پره	منفی
همراه چهارم نسبت به قیمتش میرزه و خیلی هم بد نیست	مثبت
این همراه چهارم هر هفته یه قطعی چیزی داره	منفی
نمیدونم از همراه چهارم استفاده کنم یا همراه پنجم	خنثی

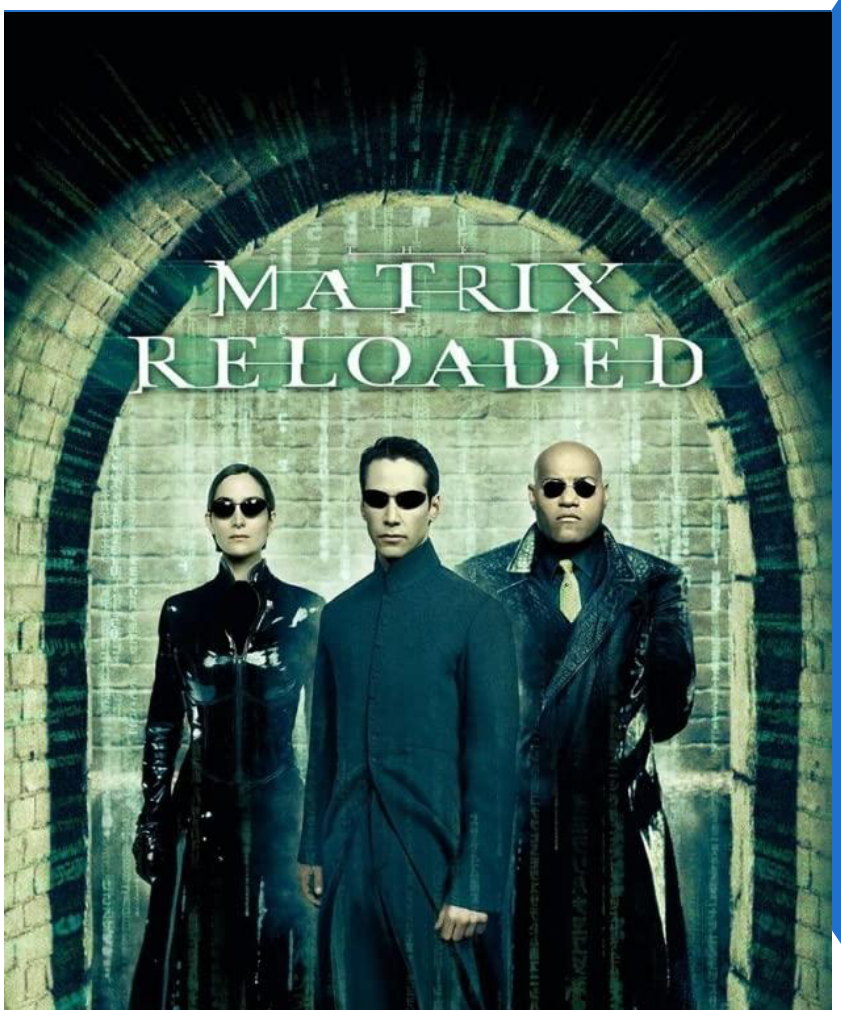
Whole Process

احساس	توییت
منفی	واقعا همراه چهارم گند زده با این خدماتش. یه روز در میون اینترنتش می‌پره
مثبت	همراه چهارم نسبت به قیمتش میرزه و خیلی هم بد نیست
منفی	این همراه چهارم هر هفته یه قطعی چیزی داره
خنثی	نمیدونم از همراه چهارم استفاده کنم یا همراه پنجم



Matrix!

Convert Data into Computer Understandable Format



Convert Text to Matrix (Feature Engineering)

- Bag of Words
- TF-IDF
- Word2Vec (Doc2Vec)
- LDA and LSI
- Glove, Bert, FastText and...

.

KhabarFarsi


استانها - پیش بینی آب و هوا - تبلیغات - صوت و ویدیو - سایت های خبری - جدیدترین خبرها - RSS
خبر فارسی
سایت هوشمند خبری
سه شنبه ۰۷ مرداد ساعت ۲۰:۴۴ به وقت تهران 〈 مصادف با 28 Jul 2020
جستجو کنید ...
خبر فارسی یک موتور جستجوگر اینترنتی است و هیچ دخالت انسانی در دریافت و چینش اخبار وجود ندارد.
خانه سیاسی ورزشی اقتصادی فرهنگ و هنر علمی و پزشکی فناوری و ارتباطات اجتماعی بین
برچسب‌های خبری
کرونا جمهوری اسلامی ایران خبر ایران سخنگوی وزارت بهداشت سخنگوی امام جمعه جمهوری اسلامی وزارت شنبه
برگزار وزارت بهداشت تیم ملی فدراسیون فوتبال

مهمترین اخبار

۲۴ ساعت گذشته | ۱۲ ساعت گذشته | ۲ ساعت گذشته | جدیدترین

واکنش نیروی دریایی آمریکا به رزمایش سپاه

برترین ها ۲ ساعت پیش + ۱۰ سایت دیگر

خبرگزاری فارس: نیروی دریایی آمریکا به برگزاری مرحله نهایی رزمایش بزرگ دریایی و موشکی پیامبر اعظم (ص) واکنش نشان داد. ...

- سرلشکر سلامی: توسعه تسلیحات ما منطبق بر شناخت نقاط ضعف و قوت دشمن است — خبرگزاری فارس / ۴ ساعت پیش
- رزمایش پیامبر اعظم 14 | تصویربرداری ماهواره نور در صدمین روز ماموریت/ نسل ... — خبرگزاری فارس / ۴ ساعت پیش

روحانی: با وثیقه سهام عدالت، کارت اعتباری بگیرید

برترین ها ۱ ساعت پیش + ۶۶ سایت دیگر

خبرگزاری فارس: حجت الاسلام حسن روحانی رئیس جمهور عصر امروز در جلسه شورای عالی فناوری اطلاعات گفت: وزرا و رؤسای دستگاه های اجرایی مسئول تحقق کامل دولت الکترونیکی هستند و شورای

جدیدترین خبرها

سیاسی

امام جمعه اصفهان: تبعیت از ولایت فقیه، وجه تمایز قوای ایرا...
ایرنا ۳ دقیقه پیش

از حربه آمریکا برای منع همکاری ایران و چین تا داستان وزیری...
باشگاه خبرنگاران ۷ دقیقه پیش

گفتمان مقاومت می تواند در شکل دهی نظام آینده جهان جای...
ایرنا ۱۵ دقیقه پیش

همکاری آمریکا با عربستان برای جاسوسی از مخالفان ریاض
ایرنا ۲۵ دقیقه پیش

ورزشی

غبیشاوری: برای اسکوچیچ در اردوی نساجی اتاق بگیرند/دوره ...
خبرگزاری فارس ۲ دقیقه پیش

توسعه ورزش آپارتمانی برنامه ریزی و در همدان دنبال می شود
خبرگزاری مهر ۴ دقیقه پیش

پرسپولیس پس از دو روز استراحت تمرین کرد
ایرنا ۱۰ دقیقه پیش

تبریک سالگرد قهرمانی آسیا با طعنه به پرسپولیس/عکس
خبر آنلاین ۱۸ دقیقه پیش

120,000

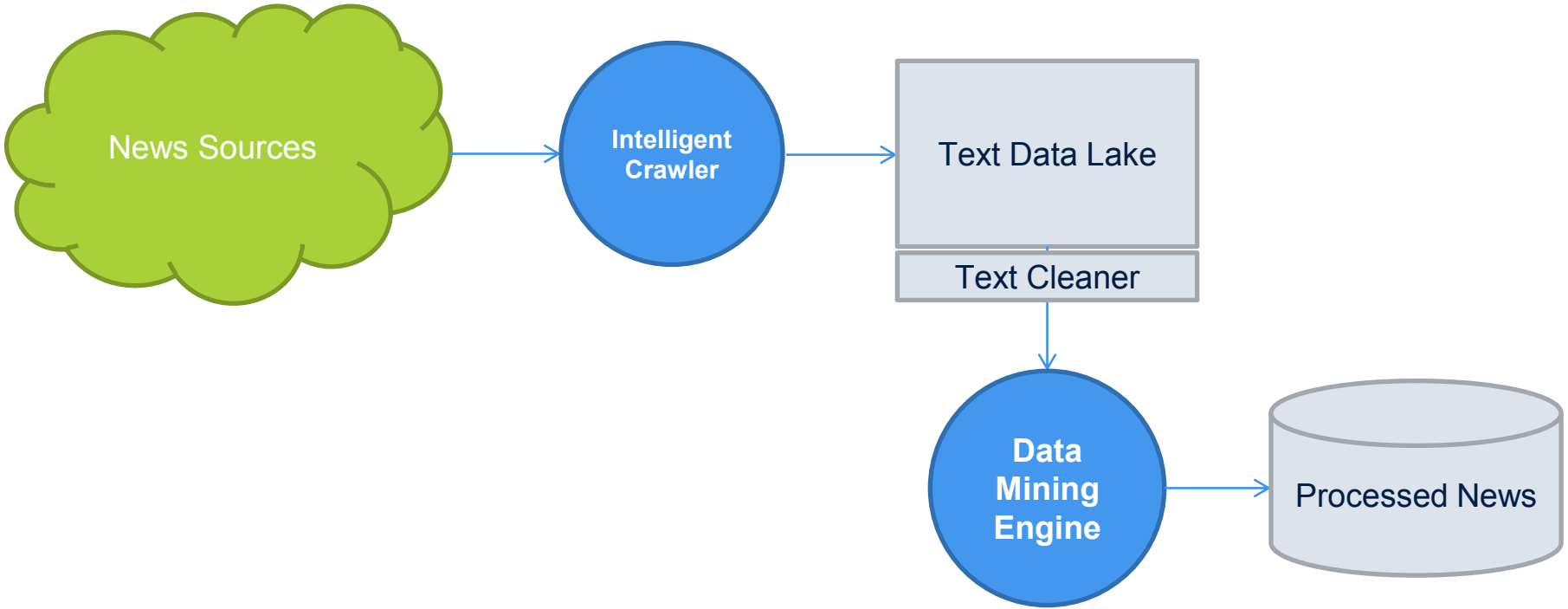
News Every Day

Glosseries

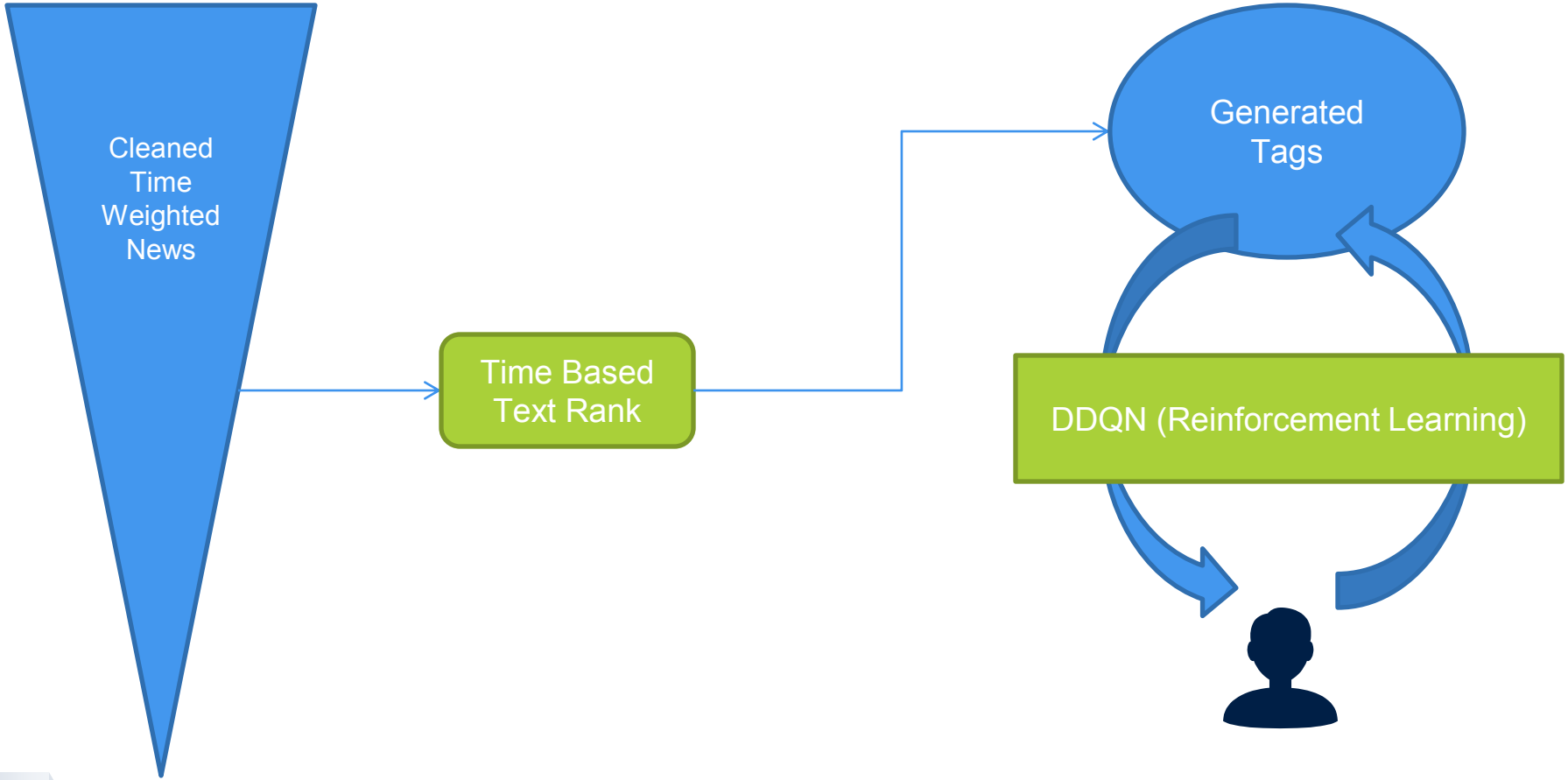
- Text Classification
- Text Clustering
- TextRank (Text Summarization)
- Topic Modeling
- Similarity Detection

.

Model 01


News Sources
Intelligent Crawler
Text Data Lake
Text Cleaner
Data Mining Engine
Processed News

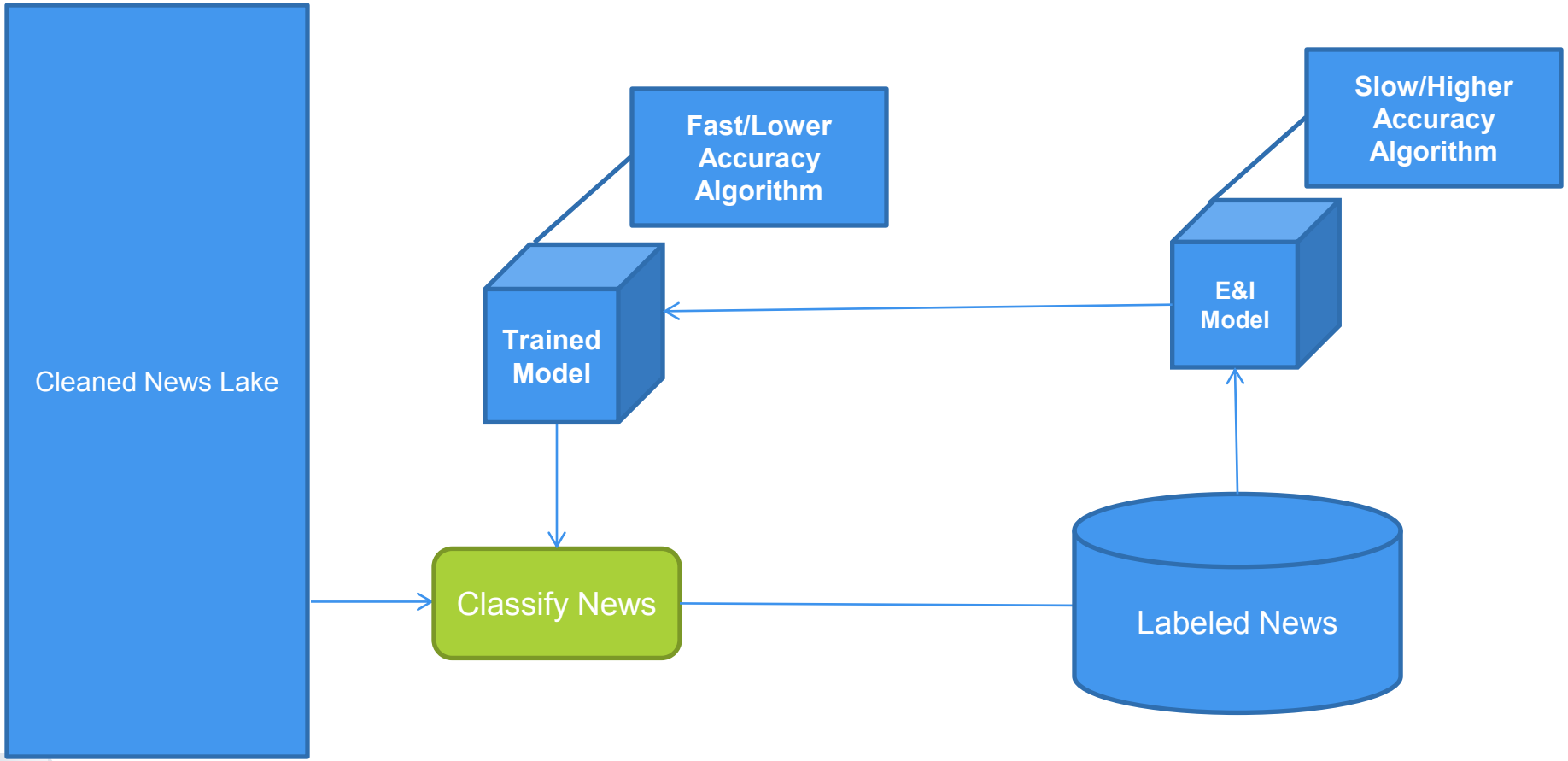
Model 02 – Popular Tags



>> Weighted Text Rank

>> Double Deep Q Reinforcement Learning

Model 03 – News Classification

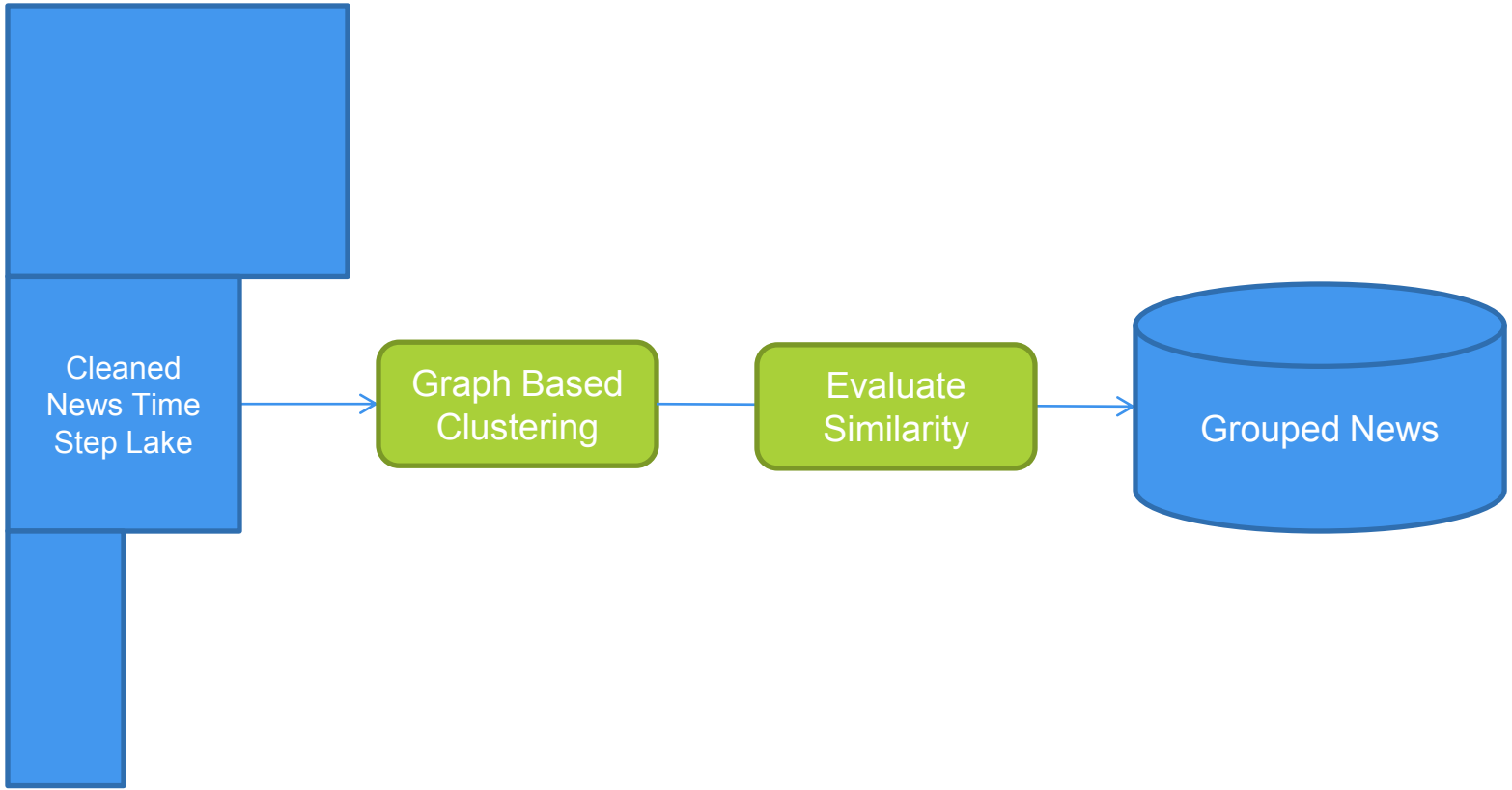


>> DNN

>> LSTM Text Classifier

>> SVM, Random Forest (With Meta Classification and Boosting Features)

Model 04 – News Clustering



>> HDBSCAN

>> Cosine & Jaccard Combined Similarity Detection

>> Spectral Clustering

Some Usefull Python Libraries

- Tensorflow, Scikit-learn,
- Gensim (Word2Vec, Doc2Vec)
- NLTK, HAZM
- CoreNLP
- spaCy

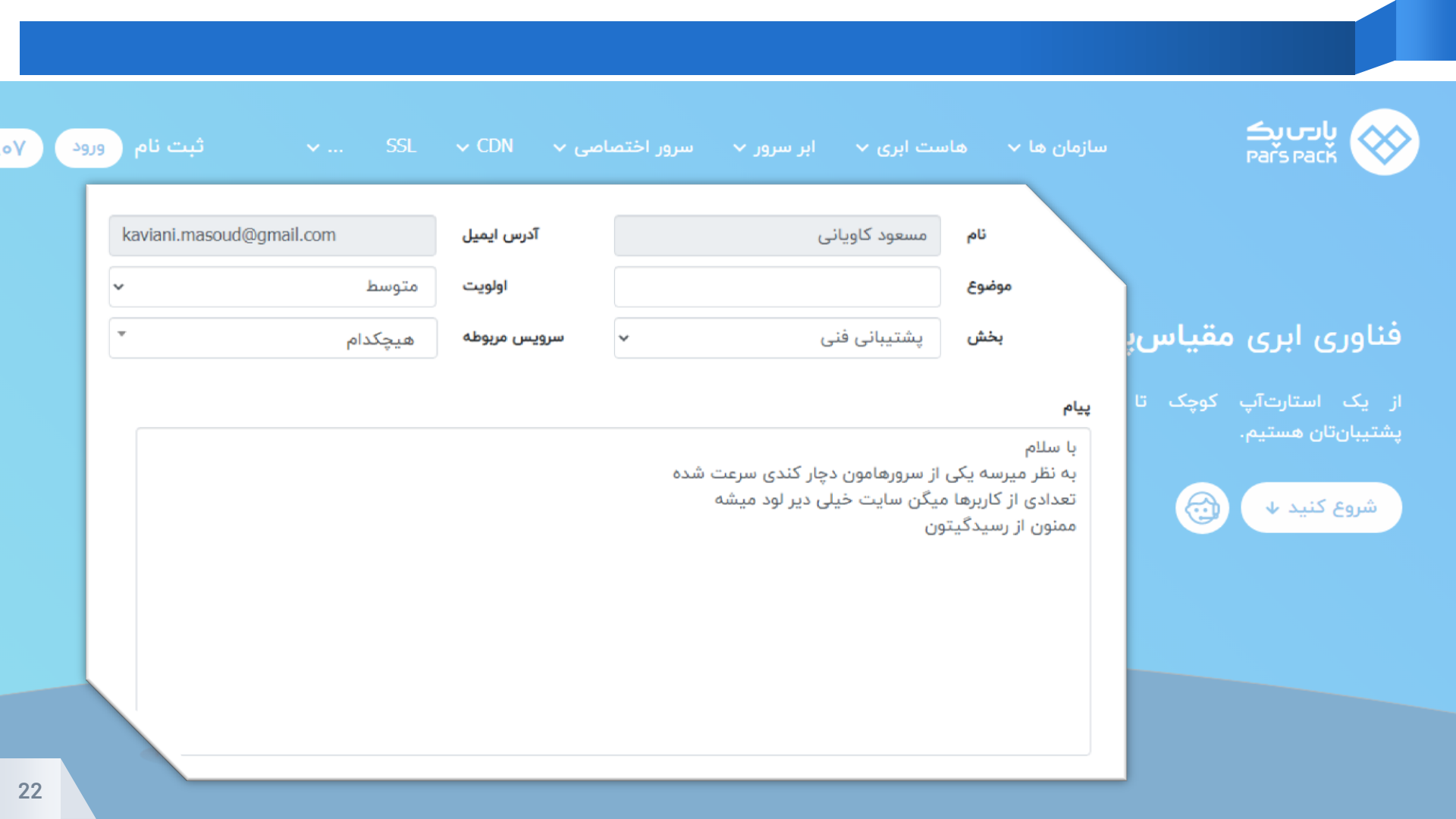
.

ParsPack

Ticketing, Log Analyter, Customer Care

Matrix Again! But...

Convert Text Data Into Matrix and Concatenate with Other Feagtures


پارس پک
Pars Pack
سازمان ها
هاست ابری
ابر سرور
سرور اختصاصی
CDN
SSL
...
ثبت نام
ورود
فناوری ابری مقیاس‌پ
از یک استارت‌آپ کوچک تا
پشتیبان‌تان هستیم.
شروع کنید
نام
مسعود کاویانی
آدرس ایمیل
kaviani.masoud@gmail.com
موضوع
بخش
پشتیبانی فنی
اولویت
متوسط
سرویس مربوطه
هیچکدام
پیام
با سلام
به نظر میرسه یکی از سرورهامون دچار کندی سرعت شده
تعدادی از کاربرها میگن سایت خیلی دیر لود میشه
ممنون از رسیدگیتون

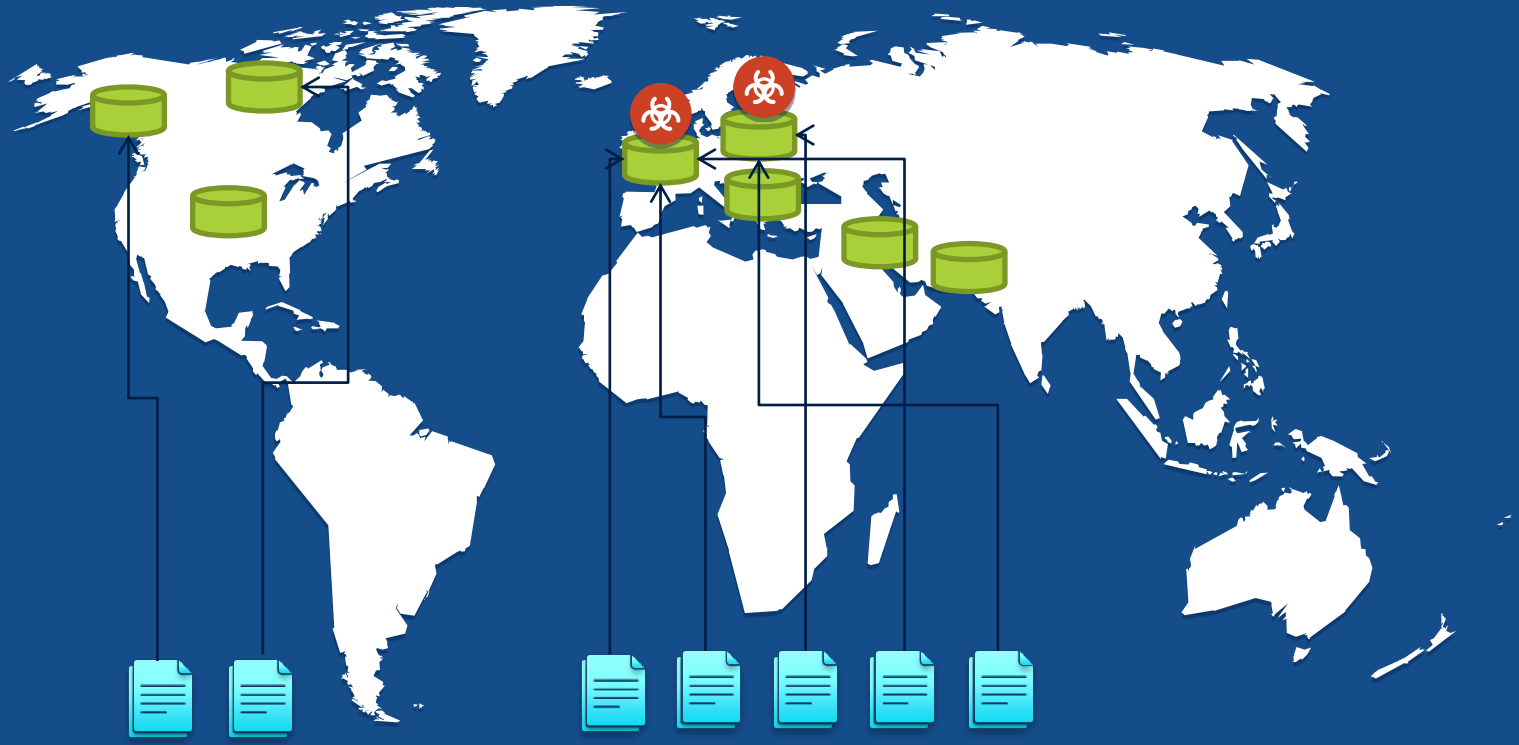
Multi Label Classification/Regression

تیکت	دپارتمان	مشکل	شدت خطا
کلا هیچی دانلود نمیشه، حتی نمی‌تونم به سرورم وصل بشم، پرداختی رو هم انجام دادم و فاکتورم پرداخت شده، ممنون میشم رسیدگی کنین	سرور دانلود	SSH, FTP	۹۰
از صبح سایت بالا میاد ولی حس می‌کنم به سرعت گذشته نیست	اشتراکی	Load	۵۷
الان با اینترنت ایرانسل دارم آپلود می‌کنم ولی حدودا یک ساعت طول کشیده	اشتراکی	Load, FTP	۴۱
من تصاویر اپلیکیشن رو بردم روی سی دی ان ولی دیباگ که می کنم ۴۰۴ میده	CDN	DNS, Ping	۸۴

What is Text Mining?

تعداد دامنه‌ها	تعداد سرورهای اختصاصی	تعداد سایت‌ها	میانگین تیکت در ماه	تیکت
۱۰	۱	۳	۳	کلا هیچی دانلود نمیشه، حتی نمی‌تونم به سرورم وصل بشم، پرداختی رو هم انجام دادم و فاکتورم پرداخت شده، ممنون میشم رسیدگی کنین
۱۲	۸	۱۳	۱	از صبح سایت بالا میاد ولی حس می‌کنم به سرعت گذشته نیست
۱	۰	۱۷	۱۲	الان با اینترنت ایرانسل دارم آپلود می‌کنم ولی حدودا یک ساعت طول کشیده
۱	۲	۸	۱۴۰	من تصاویر اپلیکیشن رو بردم روی سی دی ان ولی دیباگ که می کنم ۴۰۴ میده

Servers and Tickets



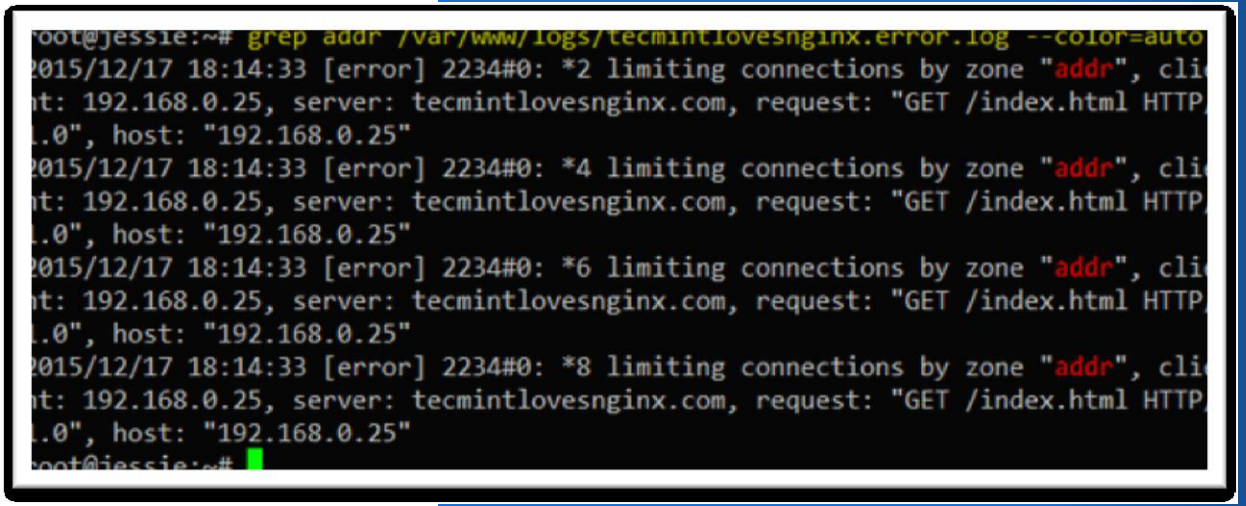
Customer Care

Prediction Customer Churn

کلمات مهم تیکت‌ها	تعداد تیکت در هفته	شیب رشد تیکت	میانگین مبلغ پرداختی در ماه	شیب رشد پرداختی در ماه	تعداد سرویس‌ها	نرخ رشد سرویس ها در ماه	ریزش
سرور – هاست – سایت – خراب – کند – فکر – پول - فاکتور	۴	۰/۴۲	۱۲۰۰۰۰	۱/۱	۱	۰/۵	۰
خراب – کمک – همکاری – هلند - کنار	۱	۲/۴۱	۴۹۰۰۰	۱	۵	۱	۱
هاست – عکس – تصویر – پینگ – ۴۰۴ – ارور	۱	۰/۴۱	۸۹۰۰۰	۰/۹	۳	۰/۴	۰
سرور – کمک – ایران - ۴۰۴	۰	۱۲/۱	۱۲۰۰۰	۱	۱	۱	۰

Log Analyzing

Logs Generate Very Fast and it is Hard to Mine these Logs at Real Time



Some Outlier Detection Algorithms

- Density-based OD
- One-Class SVM
- ISOLATION Forest
- Angle Based OD

Thanks!

Any questions?

You can find me at:

- Masoudkaviani.ir
- Chistio.ir

Image Refs:

https://storybydata.com/datacated-challenge/text-mining-for-better-insights/

https://screencrush.com/the-matrix-reloaded-15-anniversary-defense/

https://www.imdb.com/title/tt0435998/

https://www.zeliosanalytics.com/blog/how-to-predict-customer-churn

https://www.tecmint.com/nginx-web-server-security-hardening-and-performance-tips/

https://scotthelme.co.uk/monitoring-http-2-usage-in-the-wild/