Instagram Data e-mail Value Generator

Name Ivica Dimitrovski

Date 2021

E-mail contact@hpc.mk

Industrial HPC Course

SUCCESS STORY IN DIGITAL MARKETING

COMPANY: Influencers Club

https://influencers.club/
St. Venjamin Machukovski 20/1 – 15
Skopje, North Macedonia

THE PROBLEM

Increase the efficiency of the data team of the company by 100% using AI for data analytics.

SUCCESS STORY DETAILS
HPC provider FCSE, UKIM
Domain Expert Ivica Dimitrovski
Country: North Macedonia
Link:https://www.hpc.mk/index.php/events/eventhpc1/

THE HPC PROBLEM DOMAIN

Natural language processing (NLP)

Real time search, analysis and visualization of large-scale data

THE SOLUTION

The models support various NLP tasks: language identification to determine the language of text, sentiment analysis to identify positive vs. negative sentiment, Named entity recognition (NER) and text classification.

THE BENEFITS

Handle petabytes of data

High productivity

Near real time search, analytics and visualization

Improved search performance

SUCCESS STORY IN DIGITAL MARKETING

THE PROBLEM

Increase the efficiency of the data team of the company by 100% Time to Batch (Currently 72h). Things that we want to investigate here: usage of AI techniques for data analytics, improving Postgres performance, improving scraping speed, increasing the scale of our scraping power, implementing real-time collection of data, utilizing massive amounts of processing power to process and analyze real-time data at scale.

THE HPC PROBLEM DOMAIN

Natural language processing (NLP)
Real time search, analysis and visualization of large-scale data

THE SOLUTION

Introducing the ELK (Elasticsearch, Logstash, and Kibana) stack as a solution for the given problem.
Elasticsearch is a search and analytics engine. Logstash is a server-side data processing pipeline that ingests data from multiple sources simultaneously, transforms it, and then sends it to a "stash" like Elasticsearch. Kibana lets users visualize data with charts and graphs in Elasticsearch. The Elastic Stack processes data upon ingest, it uses the inference ingest processor to apply given machine learning model to the incoming data at ingest time without ever leaving Elasticsearch. The models support various natural language processing (NLP) tasks: language identification to determine the language of text, sentiment analysis to identify positive vs. negative sentiment, Named entity recognition (NER) and text classification.

THE BENEFITS

- The implemented solution can easily be scalable to handle petabytes of data both in a structured and unstructured format
- High productivity with parallel processing by allocating primary and replica shards across all available Elasticsearch nodes
- Near real time search, analytics and visualization of large-scale data (after one second the added document is searchable in this engine)
- Improved search performance
- Machine learning support - pre-trained data frame analytics model or a model deployed for natural language processing tasks to infer against the data that is being ingested in the pipeline

EURO