Zhan C.

Senior Data Scientist

Professional summary

A qualified specialist with 10+ years of experience in evidence-based decision making and data-driven development. An experience includes dealing with DWH, creation of solution architectures from scratch in cloud platforms and RESTful API design and development. Skilled in overseeing complex data pipelines which allow Customers to find out hidden behavioral patterns, identify clusters and linkages between objects of interest. Experienced in charging of evidence-based decision making and have proven leadership skill. A great team player who knows how to communicate with teammates and management. While working as a BI Team Lead was responsible for reporting for top managers. Advanced Linux user.

Education

Academy of Public Administration under the aegis of the President of the Republic of Belarus. Qualification awarded: Master of Public Administration.

Industries

Technology, Healthcare, Service Sector

Technologies and tools

- Python, R, ML
- Keras, TensorFlow, Flask, Scikit-learn, Pandas, Dask, SparkML
- BigTable, Redis, CloudSQL, MySQL, Click House, Presto, Oracle, BigQuery, PostgreSQL, Athena, Redshift
- Redash, QuickSight, Data Studio
- Docker
- Google Cloud Platform (Dataproc, Dataflow, PubSub) AWS (RDS, EMR) Azure, Yandex Cloud
- Jupyter, DataLab, PyCharm, Atom
- NLP, Statistics, Evidence Based Medicine, Bioinformatics
- Git, GitLab

PROJECTS

Water and energy industry equipment company
Duration: 07.2020 – 09.2020

With high quality measuring systems and accompanying services the company offers integrated and well-rounded solutions for the wide-ranging requirements of the water and energy sector. We are developing software platform for analysis of data on the operation of treatment facilities.

Team
4 specialists

Position
Senior Data Scientist

Responsibilities
• General solution architecture development and cost optimization
• Data pipeline development
• Anomaly detection streaming jobs development
• Database architecture
• Development of dashboards
• Team support and consulting

Technologies and tools
• Azure Cloud (storage, steaming analytics, service bus queue and topics / subscriptions, cloud SQL Postgres), Grafana, DataGrip, Git (Azure DevOps repositories)

COVID-19 determination app

Duration: 05.2020 – till now

The project is focused on creating an application, that helps to determine COVID-19 by analyzing samples of a patient's biological material that are suspected to contain viral RNA. It's mainly based on machine learning (creating algorithms, that compare results with prescribed ones and give actual result). Such system helps to give almost immediate results, saving time and patients' lives.

Team
7 specialists

Position
Senior Data Scientist

Responsibilities
- Automated interpretation of qPCR results for COVID-19 diagnostics
- Model development, evaluation and deploy, ML API development
- Research on optimization of UX\UI for laboratory employee
- Database and Data Warehouse architecture and development
- Solution and ETL pipeline architecture design
- RnD consulting

Technologies and tools
- Python, Flask, SQL, PostgreSQL, Cloud Run, Cloud Pub\Sub, Google Cloud Platform, Cloud Functions, Docker, BigQuery, Dataflow, CatBoost, PyCharm, Jupyter, Git

Survey platform

Customizable survey platform with automatic processing of results with the ability to integrate with ERP systems. The platform allows companies to conduct surveys. Survey results are processed by the ML model, which do clustering and user segmentation, reveal the importance of the function. The results are presented in dynamic and understandable information panels, separately for each client. The whole process of planning, scheduling, sending and operating the survey is carried out automatically.

Team
9 specialists

Position
Senior Data Scientist

Responsibilities
- Developed the model for customers segmentation and profiling
- Automated detection of impactful factors in survey results

Technologies and tools
- Python, Scikit-Learn, CatBoost, HDBSCAN, Amazon SageMaker, S3, SQS, Jupyter, PyCharm, GitLab

Media Service Company

Our team has built customer data pipeline, ML-based engagement rate predictions, and recommender system.

Team
7 specialists

Position
Senior Data Scientist

Responsibilities
- ML model training & deploy
- Reported for top management
- Designed cloud platform solution architecture
- Designed pipeline and data flow
- Built the DWH
- Provided mentoring

Technologies and tools
- Python, Redis, Big Query, Google Cloud Platform, Cloud Storage, Cloud Balancer, Compute engine, PubSub, Dataflow, Dataproc, Cloud Bigtable, Google ML, Google Analytics 360, MS SharePoint, Datalab, Jupyter, Git

Cloud automated reporting system

Duration: 10.2017 – 09.2018

Data from more than 50 pharmacies stored in one DWH. System daily runs queries and scripts for near 50 strategic reports and put results into columnar storage. Top managers using Quicksight or web interface to work with results, end users have pretty dashboards which also updated daily.

Team
18 specialists

Position
BI Team Lead

Responsibilities
- Predicted analytics and metrics
- Designed cloud platform solution architecture
- Created dashboards and reported to top managers
- Wrote code in Python, SQL, BASH
- Lead the team

Technologies and tools
- Python, MySQL, Pervasive, REST APIs (JSON, XML), Athena, Redshift, AWS Glue, S3, EC2, Lambda, SNS, Cloud Watch, QuickSight, Data Grip, MySQL Workbench, Jupyter, PyCharm, Pentaho, GitLab

CRM for a network of US-based pharmacies

Duration: 02.2017 – 10.2017

CRM system that helps customers to work with patients, from interest in particular medication to ship it. System covers patient's life cycle in company including following: getting documents from physicians, check eligibility, billing, shipping, provides documents to insurance companies, etc in Part B and Part D regulations. Following services were involved: FedEx, USPS, RelayHealth, Medicare, Medi-Span and others. System cover more than 3 millions patients from US.

Team
60+ specialists

Position
BI System Analyst

Responsibilities
• Reported automatization with Python and R, ETL
• Designed and architectured the database
• Wrote and optimized SQL queries

Technologies and tools
• PHP, MySQL, Pervasive, S3, Dynamo DB, MySQL Workbench, Data Grip, Pentaho, GitLab

Fundamental researches in gene expression, bioinformatics, evidence-based medicine

Duration: 09.2010 – 01.2017

Department work as service team for other scientific units. We made evidence based statistical analysis for PhDs almost in all medical directions: surgery, public health, hematology, genetics and bioinformatics, pathological and normal physiology and anatomy, obstetrics and gynecology, laboratory diagnostics and x-ray image automated analysis.

Position

Medical Researcher

Responsibilities

- Mathematical modelling
- Designed the research
- Managed the team, human resource, experiment
- Growth-hacking

Technologies and tools

- C++, SQL, R, MySQL, SPSS, Statistica, Nokia QTCreator, Excel, some specific tools for reverse transcription real-time PCR