

Ekaterina Kasilina

LinkedIn: <https://www.linkedin.com/in/katy-kasilina/>

Github: <https://katykasilina.github.io/>

Kaggle: <https://www.kaggle.com/abramova>

ek.kasilina@gmail.com +382 68 624 986

Senior Data Scientist, EPAM SYSTEMS

09.2022-now

P&G products demand forecasting for North American retailers (Costco, Sam's, Target etc)

- Developed ML models for demand forecasting with LightGBM of personal care products for the largest retailers in North America which improved MAPE of promo period planing by 23% comparing to human planing.

Responsibilities:

- features engineering and features selection
- data acquisition and data quality testing
- model tuning and validation
- writing production python code for e2e model train with pyspark and lighgbm
- pipelining models' train with Databricks workflows

Tools: Databricks, pyspark, lightgbm, azure data storage, pandas, sklearn

Out of the project activities

- Took part in creation EPAM NLP courses
- Mentoring junior and middle level data scientists

Senior Data Scientist & Team Lead Data, Inleksys

Apr 2017 - Aug 2022

Developing application for data quality and anomalies detection in telecom data.

- Developed an application for monitoring data quality and detecting anomalies in stored telecom data. This tool enhances the quality of data utilised for marketing communications. As a result, communication effectiveness surged by 15%, while complaints about poor personal offers dropped by 10%. Various models (linear regression, ensembles) and heuristics were used.

Responsibilities:

- leading ds team and the application support team (6 members in total)
- working closely with backend development team
- presenting results to stakeholders
- data exploration and feature engineering
- writing production python code for the full model lifecycle

Tools: python, sklearn, MongoDB, pandas, numpy

Developing voice bot for telecom call center

- Created a voice bot using ML algorithms that covers 24% of B2C outbound calls for telecom company which allowed the company to proportionally reduce the costs of its call center.

Responsibilities:

- leading ds team (2 members)
- exploring calls scripts, collecting data
- exploring and testing various libraries for speech-to-text, text-to-speech
- developing models

Tools: python, sklearn, pandas, numpy, yonder speech kit

Data Analyst & Team Lead, AT Consulting

Sep 2010 - Apr 2017

- Built an analytical datamart for telecom company using data from Hadoop and Oracle databases with PL/SQL
- Created analytical reports and presented them to stakeholders.
- Getting and cleaning data for machine learning models
- Created a data analytics team which had grown from 3 to 22 members.
- Created a SQL course in the company.
- Providing career mentorship and guidance

Tools: Python (pandas, sklearn, pytorch, seaborn), SQL (Oracle, Postgres), Docker, MLflow, Tableau, Hadoop, PySpark, Databricks

ML: EDA, linear models, ensembles (LightGBM, Catboost), NLP, time series

Misc: git, bash, pytest, Jira

Languages:

English: C1, French: A2, Russian: native

Education: Voronezh State University, Master Degree in Computer Science, 2010