Elizaveta Kolmakova

Data Scientist, ML engineer

Data Scientist with experience in CV, NLP and classic DS. Graduated from the Master of Science at Innopolis University with a full scholarship.

Experience with PyTorch, Tensorflow, as well as with classical DS, with basic libraries: pandas, numpy, sklearn, scipy, plotly.
Have more than a 2.5 year of commercial experience in production, where I worked with CV and NLP tasks (NER, dialog systems, NLU searching).

Most interested in projects in the field of NLP for the development of conversational AI systems or projects with the opportunity to gain experience in areas such as: voice assistant, virtual assistant, medical data processing.

☎ +7 (999) 156-03-59

📱 Telegram: @eksver

🌐 github.com/Elizaveta55

✉ e.kolmakova@innopolis.university

my post on Habr

EXPERIENCE

JANUARY 2023 — CURRENT
1 YEAR

1221Systems

DATA SCIENTIST

- Application of ML and DS techniques and algorithms on one of largest e-commerce retail mostly for recommendation systems
- Teamwork, analysis of needs and requirements at all stages of development: data analysis and setting business goals -> data collection and preprocessing -> building a solution -> analysis of metrics -> optimization -> production.

Main Projects:
1. Recommender System for HoReCa for over 52.000 clients. [knn, SVD, FastApi, YandexCloud S3, word2vec]
2. Recommender System for Retail Mobile App and Website. [ChatGPT, SVD, Streamlit, FastApi, YandexCloud S3, Implicit, LightFM]
3. Grocery description generation for website. [YandexGPT]
4. Intent Classifier for customer support [LaBSE, Bert Classifier, TelegramAPI]

Achievements:
1. Over 90.000$ additional profit after launching Personal Recommender System for HoReCa
2. Gain expert recommendation without inviting third parties.
3. High quality of generated text suitable for website description using ChatGPT and YandexGPT.
4. Build a classifier for Russian without domain russian dataset.

DECEMBER 2021 — JANUARY 2023
1 YEAR 1 MONTH

ZeBrains

DATA SCIENTIST

- Application of ML and DS techniques and algorithms on real problems to solve business needs.
- Teamwork, analysis of needs and requirements at all stages of development: data analysis -> preprocessing -> building a solution -> analysis of results -> optimization -> production.

Main Projects:
1. Application of regression models for the largest medical customer to predict analysis scores. [Sklearn, regression models, gradient boosting regression, statistics]
2. CV detection, classification and object type recognition on the camera. [PyTorch, U2net, OpenCV, yolo]
3. Chatbot with NLP recommending system for films in Russian, based on text analysis of movie descriptions parsed from Wikipedia.
NER recognition was added for more accurate recommendation. [tensorflow, HNSW, LaBSE, natasha, spacy, DeepPavlov, ELMo]

Achievements:
1. Developed working regression models without overtraining in terms of highly limited dataset. Obtained the desired level of reliability required for the medical domain.
2. Developed an algorithm for generating data in terms of limited dataset: implemented the functionality of adding new classes to the dataset.
3. Implemented a fast contextual search engine for a large number of films (43 thousand films) with average searching time 0.00144s

OCTOBER 2021 — OCTOBER 2022
1 YEAR 1 MONTH

YS Vision

DATA SCIENTIST

The startup provides an AI assistant for the young generation to support their potential and well-being.
- Advising and implementing AI applications, starting with a intent classifier.
Achievements: rose to the position of one of the core team for the ideation and development of the product.

MAY 2021 — JUNE 2021
2 MONTHS

EORA data lab

DATA SCIENTIST

Development and implementation of a conversational AI chat-bot based on two approaches: generative (ruGPT) and retieval-based, plus subsequent ranking of the results. DeepPavlov is used for ranking.
- Collecting a dataset, setting up models, analyzing results, deploying a service (FastApi).
[NLP, ruGPT, DeepPavlov, FastApi]

UNIVERSITY

2020-2022

Master Science

INNOPOLIS UNIVERSITY

- Multiple team projects and hackathons: Inno Science Hack (1st place), Management Hackathon Use in tools (2nd place).
- Completed courses: Machine learning, Advanced Machine learning, Advanced Information Retrieval, Big Data, Managing Software Development, Optimization, Advanced Statistic, Empirical methods.
- Research and implementation of master thesis "When and Why Generative Models (GAN) are Useful for Imbalanced Classification Problems?"
[pytorch, GAN+AE, parameters selection, unfrozen weights]

2016-2020

Bachelor

KAZAN FEDERAL UNIVERSITY

- Project #1: Application of Neural Networks to Predict Sentiment of a Message [Python, tensorflow, LSTM].
- Project #2: Application of Neural Networks to predict the treatment status of patients with myocardial infarction. [Python, tensorflow, LSTM, word2Vec, data preprocessing]