Murat APISHEV

Moscow / Russia | +7-916-769-82-75

mel-lain@yandex.ru

github | My page (in russian)

SUMMARY

I am an expert in data science, machine learning, natural language processing and software development. My skills include data analysis, training ML and DL models, designing and writing efficient code, team leading. Also I teach a lot and like to be a public speaker.

EDUCATION

MOSCOW STATE UNIVERSITY (CMC) / FRC IM OF THE RUSSIAN ACADEMY OF SCIENCES   *2017-2020*

PH.D IN COMPUTER SCIENCE

•Effective Implementation of Topic Modeling Algorithms with Additive Regularization

MOSCOW STATE UNIVERSITY (CMC)   *2015-2017*

MASTER OF APPLIED MATHEMATICS AND COMPUTER SCIENCE

MOSCOW STATE UNIVERSITY (CMC)   *2011-2015*

BACHELOR OF APPLIED MATHEMATICS AND COMPUTER SCIENCE

WORK EXPERIENCE AND PROJECTS

TALKMART42   *April 2020 - Now*

DATA SCIENTIST (CO-FOUNDER)

•Prototype of voice control system for self-service checkout for retail company *(Python / NLP)*
•English neural QA-system for specific domain for advertising agency *(Python / NLP)*
•Advertisements classification system for advertising and media research company *(Python / NLP)*
•Voice games for Yandex Alice and Sber Salut *(Python)*

DIGITAL DECISIONS (AITHEA)   *April 2019 - April 2020*

NLP TEAM LEADER / DATA SCIENTIST (SENIOR)

•Spell checker for government organization *(C++ / Python / NLP)*
•English NER in application onboarding for B2C startup *(Project managment)*
•Search and recommender system for English scientific articles *(Project managment / Team leading)*
•Trend detection in Twitter and YouTube data for Instinct BBDO *(Python / NLP / Team leading)*
•Collecting NLP-data for Huawei voice assistant *(Python / NLP / Team leading)*
•Tickets ordering chatbot for B2C company *(NLP / Team leading / Project managment)*

YANDEX ZEN   *September 2017 - April 2019*

ML-ENGINEER (MIDDLE)

•System for evergreen content detection within Zen recsys *(Java / Python / SQL / ML)*
•System for one-day/yellow/black content classification within Zen recsys *(Java / Python / NLP)*
•Zen recsys channel catalog ranking system *(Java / SQL)*
•Zen recsys fresh content usage strategies algorithms *(Java / SQL)*

YANDEX SEARCH   *April 2016 - September 2017*

SOFTWARE ENGINEER (JUNIOR)

•Yandex Search antirobot development *(C++)*

COMPUTING CENTRE OF RUSSIAN ACADEMY OF SCIENCES   *July 2014 - April 2016*

SOFTWARE ENGINEER

•BigARTM - open-source library for topic modeling *(C++ / Python / NLP)*
•Ethnic content exploration in social networks data *(Python / NLP)*
•Clustering and sentiment analysis of bank employees reviews *(Python / NLP)*

MAIN TEACHING EXPERIENCE

MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY   *2019-2021*

LECTURER OF PYTHON COURSE

HIGHER SCHOOL OF ECONOMICS   *2020*

LECTURER OF NLP ONLINE COURSE

Moscow Institute of Physics and Technology / Moscow State University   *2017-2020*
Lecturer of NLP Course (earlier - Assistant)

Yandex School of Data Analysis   *2016-2019*
Assistant in ML Course

Higher School of Economics   *2017-2018*
Assistant in NLP Course

Skills, Models, Approaches

Programming Languages	*Experienced:* Python 2/3 | C++11 ***Familiar:*** Java | JavaScript | SQL
Software Development	Algs and Data Structures | OOP Patterns | Multi-Threaded Programming | MapReduce
General ML and DS	Log/Linear Regression | kNN | RF | GB | PCA | t-SNE | K-means | GMM | DBSCAN | SMOTE | Time-Series Exp. Smoothing | ARIMA | A/B Tests | Stat. Hypothesis Testing
Non-deep NLP	Text Preprocessing | Regexes | TF-IDF | Word2vec | GloVe | FastText | LDA | ARTM
Deep NLP	FFNN | CNN | GAN | RNN (BiLSTM) | Seq-to-seq (Attention) | Transformer | BERT | RoBERTa | ALBERT | ERNIE | DistillBERT | XLM | GPT-2 | XLNet | T5 | BART
Technical Skills	Linux Terminal (Ubuntu) | VC (Git, SVN) | Unit Testing | CI | LaTex
Soft/Managment Skills	Teaching | Agile (Scrum, Kanban) | Waterfall | Fluent Speech (Russian)
Languages	Russian (Native) | English (Intermediate)

Tools, Libraries, Frameworks

Software Development	*Experienced:* STL | Venv | Pipenv | Docker | Redis | Flask | Google Protobuf
	Familiar: Hadoop | PySpark | MPI | CUDA | Cython | Numba | Android SDK (Java) | QT | React | GDB | CMake | MongoDB | Travis CI | AWS
General ML	NumPy | Pandas | Sklearn | Matplotlib | SciPy | XGBoost | CatBoost | Vowpal Wabbit | Statsmodels | Faiss | ScaNN | Jupyter Notebook | Colab | Yandex Toloka
Non-Deep NLP	Re | Pymorphy2/Pymystem3 | NLTK | SpaCy | FastText | Gensim | BigARTM | UDPipe
Deep NLP	PyTorch | Huggingface | Sentence Transformers | Haystack | DrQA | EasyNMT
Voice Assistants, ASR/TTS	Yandex SpeechKit | Vosk (Kaldi) | Yandex Dialogs | Google Actions | RHVoice
Managment/Others	Yandex Tracker | Emacs | MS Office | Google Docs | Overleaf | Beamer

Main public activities

- Alice Urban Skills Hackathon, 2021 (with team, 3rd place)
- VirusHack Hackathon, 2020 (with team, 1st place)
- DataStart Conference, 2020 (two paired speeches, Instinct BBDO and Huawei)
- OpenTalks.AI Conference, 2020 (paired speech, Huawei)

Main scientific papers

- **Apishev M.**, Vorontsov K. Learning Topic Models With Arbitrary Loss. FRUCT, 2020
- **Apishev M.**, Koltcov S., Koltsova O., Nikolenko S., Vorontsov K. Additive Regularization for Topic Modeling in Socio-logical Studies of User-Generated Texts. MICAI, 2016
- Frei O., **Apishev M.** Parallel Non-blocking Deterministic Algorithm for Online Topic Modeling. AIST, 2016
- Vorontsov K., Frei O., **Apishev. M.**, Romov P., Suvorova M., Yanina A. Non-Bayesian Additive Regularization for Multimodal Topic Modeling of Large Collections. CIKM Workshop, 2015