



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*)
- Voice and visual business apps in Sber Salut (*TypeScript / Python*)

JUST AI October 2021 - Now

NLP R&D TEAM LEADER

- NLP core of cloud text processing engine for internal products and external marketplace (*Team leading / Project management / NLP / Python*)

ALFA BANK August 2021 - October 2021

DATA SCIENTIST (LEAD)

- ML and DL models for chat-bot intent detection (*Python / NLP*)
- Project management of several NLP tasks with reporting to senior officials (*Team leading / Project management*)

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 management*)
- Search and recommender system for English scientific articles (*Project management / 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 management*)

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++*)

SOFTWARE ENGINEER

- BigARTM - open-source library for topic modeling (C++ / Python / NLP)
- Ethnic content exploration in social networks data (Python / NLP)

FREELANCE

DATA SCIENTIST

- Clustering and sentiment analysis of bank employees reviews (Python / NLP)
- Retail data analysis and parsing for scientific nutritional project (Python / NLP)
- Development of recommender system for online food store (Python)

MAIN TEACHING EXPERIENCE

ONLINE EDUCATION COMPANY NETOLOGY	2022
<u>LECTURER OF NLP / DL, ML / RECSYS AND PYTHON COURSES</u>	
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY	2019-2022
<u>LECTURER OF PYTHON COURSE</u>	
HIGHER SCHOOL OF ECONOMICS	2020
<u>LECTURER OF NLP ONLINE COURSE</u>	
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY / MOSCOW STATE UNIVERSITY	2017-2022
<u>LECTURER OF NLP COURSE (EARLIER - ASSISTANT)</u>	
YANDEX SCHOOL OF DATA ANALYSIS	2016-2019
<u>ASSISTANT IN ML COURSE</u>	
HIGHER SCHOOL OF ECONOMICS	2017-2018
<u>ASSISTANT IN NLP COURSE</u>	

SKILLS, MODELS, APPROACHES

PROGRAMMING LANGUAGES	<i>Experienced:</i> Python 2/3 C++11 <i>Familiar:</i> TypeScript Java 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 DistilBERT XLM GPT-2 XLNet T5 BART
TECHNICAL SKILLS	Linux Terminal (Ubuntu) VC (Git, SVN) Unit Testing CI LaTeX
SOFT/MANAGEMENT SKILLS	Teaching Agile (Scrum, Kanban) Waterfall Fluent Speech (Russian)
LANGUAGES	Russian (Native) English (Intermediate)

TOOLS, LIBRARIES, FRAMEWORKS

SOFTWARE DEVELOPMENT	<i>Experienced:</i> STL Pyenv & Venv Pipenv Docker Redis Flask Google Protobuf <i>Familiar:</i> 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 Doccano
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
MANAGEMENT/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

- Artemova E., **Apishev M.**, Sarkisyan V., Aksenov S., Kirjanov D., Serikov O. Teaching a Massive Open Online Course on Natural Language Processing. NAACL Workshop, 2021
- **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 Sociological 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