ANDREW CHAUZOV
Data Scientist
Machine Learning Engineer
France

✉ **avchauzov@gmail.com**

in **/avchauzov**

⚙ **/avchauzov**

– Boasting over 10 years of experience in Data Science & Analytics, with 7 years in Python & Machine Learning and 5 years in Deep Learning & Engineering.

– Recently focused on Natural Language Processing, showcasing skills in text classification, topic modeling, Transformers, LLMs, GenAI Interaction.

– Strong foundation in machine learning methodologies and statistics.

– SQL, Big Data, API development and testing, and deploying solutions on GCP/AWS.

– Proficient in communicating complex data insights and mentoring.

Experience

03/2020 – Present: ML Engineer & Data Scientist at **TOPTAL**
A digital platform connecting skilled talent with global business opportunities.

– Formulated a BERT-based approach for Instagram profile categorization, achieving over **80% accuracy across 50+ groups** and **enhancing ad campaigns**.
– Employed a nearest neighbors model for **Twitter account category estimation (~4*16 subclasses)** using TFHub pre-trained embeddings.
– Implemented audience enrichment with hypothesis testing and SQL, generating over **100k potential customer matches per company**.
– Engineered over **15 predictive models for 100+ million IDs** via GCP using PL/SQL and regex, ensuring scalability and efficient logging.
– Leveraged OpenAI API for data generation, **standardizing job titles and estimating seniority/department with ~90% accuracy**.
– Applied random walk embeddings, UMAP, and HDBSCAN to customer interests, **enriching a targeted ad subset by 25k users**.
– Designed a Scikit-Learn **salary prediction pipeline with less than 10% error** using unsupervised transformation.

11/2020 – 10/2023: Data Scientist & ML Engineer at **CRED INVESTMENTS**
A platform for investing in and nurturing emerging talent.

– Combined classification analysis and tree embeddings for in-game location data, **identifying player roles with ~90% accuracy** and improving clustering.
– Assembled a 'team profile' algorithm using gradient boosting and statistics, **achieving 85% accuracy in pinpointing team weaknesses** within a league.
– Deployed over **10 predictive models** via REST API, **managing over 100k predictions** with rapid response times and robust data issue resolution.
– Designed a dual-layer regression model for football market value forecasts, **achieving short- and long-term predictions with <10% error**.
– **Collaborated on a 25-page Streamlit dashboard** featuring Plotly, Seaborn, and Matplotlib visuals, contributing to GitHub code reviews.
– **Created a 4-step data engineering pipeline** for player stats (cleaning, scaling, imputation), **enhancing models' accuracies by ~75%**.
– Architected a LightGBM **model with a 15% error rate**, identifying promising young talents and suggesting optimal replacements.

01/2017 – 03/2020: Data Scientist & ML Engineer at **FREELANCE**

– Applied NumPy to custom sale volumes clustering algorithm, **analyzing over 10k time series at once**, resulting in up to a **15% increase in sales effectiveness**.
– Trained a CatBoost model for **employee churn prediction task with 85% accuracy**, providing HR with actionable insights using the LIME library.

- Achieved **70-95% accuracy in detecting speech defects in children**; deployed the model for a mobile app using FastAPI and Docker.

02/2012 – 12/2016: (Senior) Data Analyst at **ASSOCIATION 'NON-PROFIT MARKET COUNCIL'**
Central hub fostering collaboration and trade in energy markets.

- Enhanced power price/volume forecasting models using feature engineering, achieving up to a **2.5x reduction in error rates** for targeted regions.
- Conducted ARIMA-based time series analysis for anomaly detection, leading to a **50% increase in prediction accuracy** and raised data quality.
- **Led 4 commercial predictive projects with teams of 2-4**, deploying models that **enhanced client revenues by approximately 10%**.

07/2011 – 02/2012: Junior Data Analyst (Marketing Mix Modeling) at **BBDO GROUP**
Top-tier network for creative marketing and advertising.

- Implemented an anomaly detection function, leading to a **50% reduction in expenditure** and improved client marketing budget efficiency.
- Introduced automated scripts, **increasing departmental task efficiency 4x** through streamlined media data collection and processing.
- Initiated cluster analysis to estimate pre-campaign efficiency, leading to more **effective budget allocation across over 10 campaigns**.

Education
Peoples' Friendship University of Russia
Master of Science - MS - Applied Mathematics & Computer Science

Projects
07/2021 – 07/2023: ML Development & Engineering for Football Scouting Recommendation System.

- Developed an algorithm to estimate player-team compatibility scores using advanced analytical techniques, including data wrangling, LightGBM models with SHAP-based feature selection, and **optimized the SQL prediction pipeline with 10x efficiency, achieving 5-15% error rates**.

Personal Projects
10/2023 – Present: AutoML: Unsupervised Model Training with Optuna & SHAP Feature Selection.

- Engineered an AutoML solution integrating best practices, including improved Optuna hyperparameter optimization, custom feature selection, and pytest for efficient development, **supporting 5 models with ongoing enhancements**.

Skills
- Pandas, SciPy, PyTorch, TensorFlow, Keras, NLP (Large Language Models, GenAI APIs, spaCy, NLTK, Gensim), Dimensionality Reduction (PCA, UMAP), TsFresh, Scraping (Selenium), Git, Jupyter, Optimisation Techniques (Optuna), Model Selection.
- Exploratory Data Analysis (EDA), Data Analytics, Data Visualization, Data Drift, Predictive Analytics, Data Mining, Data Modeling, Model Development.
- Statistical Analysis, Bayesian Statistics, Probability Theory.
- Model Deployment (Luigi, MLflow, Neptune.ai), API Testing, Software Engineering Practices.
- SQL Databases (BigQuery, MySQL, PostgreSQL), NoSQL (Redis), Cloud Services (Google Cloud Platform, Vertex AI, Amazon Web Services), Linux (shell scripting, Ubuntu, Debian).
- Network Analysis, Graph Theory.
- Stakeholders Interaction, Reporting, Agile (Scrum), Version Control.