Pavel Chernov

Data Scientist / ML Engineer

diovisgood.github.io
pavel.b.chernov@gmail.com
+7-90-4810-4810

Summary

Master of Computer Science (graduated with honors). Has 15+ years of programming experience, 5+ years as a team leader, and 5+ years in Data Science & Research. Started in different software and system integrator companies as a backend developer, system architect, and team leader. Went to develop management skills as an executive director at a banking equipment service company. Started a company for personalized laser-engraved souvenirs. After that, for more than 10 years, has been running his own business in development and real estate management. Then went back to the start and since 2019, has been successfully building a career in machine learning and data science. Has completed multiple projects in CV, NLP and data analysis. Has participated in many presale negotiations with potential clients. Personal business background helps to better understand business needs and find effective solutions.

Skills

Technical Skills: ML, DL, NLP, CV, RL, GANs, Diffusion

Instrument Skills: PyTorch, Transformers, Pandas, OpenCV, CatBoost, Matplotlib, Python, AWS, Docker

Education

Computer Science in South Ural State University, Russia, GPA: 4.85 out of 5, graduated with honors: June 2002
Online Coursera courses:

Course name	Date		Course name	Date
Game Theory	Sep 2013		Interest Rate Models	Oct 2020
Machine Learning	Nov 2013		Trading Strategies	Dec 2021
Neural Networks and Deep Learning	Sep 2017		Natural Language Processing with Probabilistic Models	Jan 2022
Improving Deep NN: Hyper-parameter tuning, Regularization and Optimization	Sep 2017		Natural Language Processing with Classification and Vector Spaces	Jan 2022
Structuring Machine Learning Projects	Oct 2017		Natural Language Processing with Sequence Models	Jan 2022
Convolutional Neural Networks	Nov 2017		Natural Language Processing with Attention Models	Feb 2022
Sequence Models	Feb 2018		Build Basic Generative Adversarial Networks (GANs)	Feb 2022
Practical Reinforcement Learning	Dec 2019		AWS Certified Cloud Practitioner	Feb 2023

Experience

InData Labs (company specializing in AI), data scientist / ML engineer   Jan 2022 – now
Python, PyTorch, Transformers, Rust, OpenCV, ONNX, AWS
I have completed several projects, and participated in many presale negotiations. Here are only some of projects:

Project: Webpages Semantic Analysis for Ads Campaigns
Project role: Team Lead, Data Scientist, System Architect
In the industry of online ads agents need to quickly decide what ads to display on a particular webpage for a particular user, taking into account the content of a webpage and user interests. First I solely implemented and proved the idea during the PoC phase. Then, during the second phase me and a team of other engineers created MVP. I researched different approaches for web-scraping and semantic analysis, investigated the feasibility of small response timings. The challenge was to deploy this system on the AWS cloud platform. I designed the loosely coupled architecture of the whole system and implemented a part of it related to the processing of large amount of data. Also I deployed our models in AWS SageMaker.

Project: Discovering Outliers in Financial Report of Hotels
Project role: Business Analyst, Data Scientist
A platform which allows private investors to buy and sell shares in real estate business (hotels, apartments) is collecting a lot of financial reports of hotels monthly. The client had no clear understanding how to process this data, but wanted to offer some AI features to its customers. I studied available data for ~ 1.5 years, analyzed variables and selected which variables should be considered as "input" and which – "target". Target variables are mainly expenses

governed by hotel management, as we want to monitor these and get alerts for unreasonably large expenses. I carefully chose learning algorithm to avoid overfitting, trained prediction models and created a function to highlight cases, when predicted costs differ from actual ones.

Project: Bookie Platform Churn Rate Reduction
Project role: Business Analyst, Data Scientist
A specialized platform which helps bookies to run their business is facing some client churn and wants to reduce it. Initially the client wanted just a simple churn-prediction model. Instead I dived deep into its business and offered solutions of what should be changed in the work with its customers so that they would get more profits and not abandon their bookie business, thus bringing more income to the client.
- deep-diving into the bookie business to understand the workflow of their customers - bookies
- choosing metrics to measure the performance of bookies in different areas of a bookie business
- specifying the ways to compute these metrics and what source information is needed for this
- specifying the ways to help underperforming clients to increase their profits, thus reducing the overall client churn rate

Project: Exploring Trading Strategies for Crypto Assets
Project role: Team Leader, Software Architect, Data Scientist
A client wanted to perform research if it is possible to use statistics and machine learning to find some profitable trading algorithm for a crypto asset. First I constructed a list of all possible kinds of input information which could potentially be useful for the task: S&P500 Index, DXY dollar Index, dates of reports publications, various blockchain statistics: number of sending and receiving addresses, average transaction size, average commission size, etc. Then we did our best to find the data sources to harvest information from. After that we performed research to test various hypotheses.
- creating connectors to various data providers
- creating a data-loader tool, which loads and combines heterogeneous data streams into single data-frame
- creating a fast and unified backtest module to emulate exchange environment and compute performance metrics
- trying out both supervised and unsupervised learning algorithms for the analysis

Project: Sales Analysis for Food-Retail Business
Project role: Business Analyst, Data Scientist
An established Japanese food-retail business requested some simple sales analysis by geography and time. I performed the analysis and produced neat charts of how sales are distributed geographically and in time. For this I created code to convert Japanese addresses specified in Kanji, Hiragana or Katakana into geographic coordinates, code to draw heat maps showing the distribution of orders on the map.

MapHub (services for aerial footage), ML engineer   Sep 2021 – Oct 2021
Python, PyTorch, OpenCV, ONNX, Docker, Win, Linux
This was a contract for a single project. I created an object detection model to detect buildings, coconut palms, etc. in GeoTIFF files. I've chosen YOLOv5 architecture as it is fast and accurate enough. First challenge was that GeoTIFF files are typically very huge. I found a lightweight library, which is capable of loading some patches from large GeoTIFF files. Second challenge was that inference code from YOLOv5 is not capable of processing large GeoTIFF files by parts. That is why I created my own code, which runs model on multiple patches of source image and implements non-maximum suppression algorithm to produce the final predictions list. Third challenge was to build a docker image of some reasonable size. By converting model from PyTorch to ONNX format, I managed to create a lightweight docker image of only 300 Mb.

RTS Munity s.a. (e-sport odds and trading provider), senior data scientist   December 2019 – Oct 2020
Python, R-language, PyTorch, Tensorflow, OpenCV, Win, Linux, AWS
I performed several projects for this company.

Project: Real-time OCR Reader
Project role: Data Scientist
I managed to create a real-time OCR reader of different values of teams and players on video streams of League of Legends games. The first challenge was that there are 57 different values, such as: gold, kills, turrets, minions, health, mana, etc., which can be located in different places for different video stream providers. My code automatically recognizes layout (coordinates) of these values. The second challenge was to read all these values in real-time. I carefully constructed CRNN -based model architecture, to reach an optimal balance between speed and accuracy.
Examples: [video1](), [video2](), [video3](), [video4](), [video5]()

Project: Analysis of Blockchain Betting Platforms
Project Role: Business Analyst

I've been asked to analyze new blockchain betting technology. I studied and described existing kinds of blockchain bets: Pool bets, Peer-to-peer bets, Multi-peer bets. I performed analysis of about 18 past and present blockchain projects, specially designed for making bets. I researched when each project was created, how it developed and where it is going today, or has already been abandoned.

Project: Pre-match prediction models
Project role: Data Scientist
Implemented a bundle of small models for predictions of different match outcomes for bets on e-sports games. Predictions were made based on some preliminary data available. The challenge for me was the necessity to use the R-language, which I had no experience earlier, and creating a complete R-package with unit-tests and documentation.

Osnova Ltd. (real estate management), deputy director, co-founder   July 2010 – December 2021
In addition to daily work I developed and implemented software (**Excel + VBA**) for diverse investments analysis which takes into account cash flow and calculates annualized rate of return. This makes it easy to compare results of investments in completely different kinds of instruments, for instance: commercial real estate vs stocks or bonds.

Laserographic Studio Ltd. (personalized laser engraved souvenirs), Founder   Nov 2008 – June 2010
Besides creating and running new business I adapted existing basic accounting software (1C) for souvenir production.

Pallada Ltd. (local dealer of banking equipment), executive director   Sep 2006 – Nov 2008
Besides my daily work:
- Programmed additional module for working software (**1C**) to speed up manual operations of sales managers.
- Adapted existing basic accounting software (**1C**) to work with many branches in the distributed database mode.
- Developed and realized new scheme for Internet access for all branches to make internal network and cut costs.
- Developed and created system to control Internet access for workers as it was expensive at that time.

Astra-ST CJSC (system integrator), team leader, **Linux C/C++**   Sep 2002 – Sep 2006
My team created and certified special Linux distributive with safe boot, additional access control by different security levels, with server and clients working in graphical terminal mode. This project required wide variety of skills including: creating boot images and kernel module, adapting network protocols, modifying print stream on-the-fly, creating new graphical interfaces (C/C++, Assembler, Bash, PostScript).

DataQSoft (full stack software company), backend and desktop developer, **Win C/C++**   Feb 2001 – June 2002
Created module for speech compression for voice chat over Internet (C/C++).
Created client/server for TCP tunnel over UDP protocol to connect two branches through restricted network (C/C++).
Created backend server module for online textual web-chat (C++).

Link-Service Ltd. (accounting software company), developer, **Win 1C and C**   June 2000 – Sep 2000
Created module for sales software to connect electronic balance to speed up sales and reduce operator errors.

Personal Research Projects

Reinforce – deep study of reinforcement learning (Pytorch, Python, RL)
I studied several reinforcement learning algorithms and implemented them in code from scratch. The algorithms are:
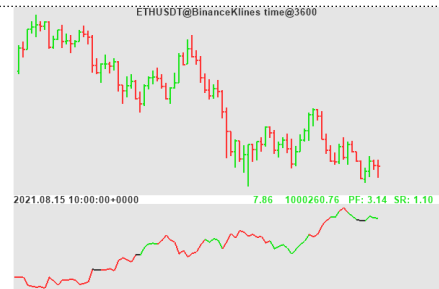- CMA-ES
- DQN (+Double, +Duelling)
- A3C
- A2C
- PPO
  https://github.com/diovisgood/reinforce

Intraday – Gym environment for intraday trading bot (Python, Numpy)
I created a package, which provides gym compatible environment to simulate intraday trading based on stream of trades, either historical or real-time.
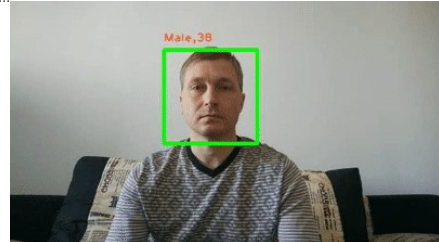https://github.com/diovisgood/intraday

Agender - age and gender real-time estimation (Python, TensorFlow, OpenCV)
https://github.com/diovisgood/agender
This is a demo project I made to learn OpenCV. It uses pre-trained lightweight models and can estimate age and gender in real time. Article on Medium.

Index - stock index price prediction (Python, Pytorch)
https://github.com/diovisgood/index

This is a demo project in which I show that you may train a good model for linear regression, but it won't help you to beat the market.

QGEN – Competing Genetic Algorithm to find profitable trading strategies in financial market (Lua, Torch)
https://github.com/diovisgood/qgen_lua
This is a project I made trying new approach to find profitable trading strategies. I could not beat the market, but developed Competing Genetic Algorithm that can be useful for many other problems. Article on Medium.

Example of evolved trading strategy code in synthetic language →
with profit factor: 1.99, max drawdown: 1.4%

pc Val po Val Sub I time Val I vol Delta vwap Val 0.6506024096 Gt Mul ptr Val 0.41427952546279 Gt time Val I I I I I vwap Rank ptr Val Sub Days po Val pc Val Sub I I Days po Val time Val 0.32497333249298 Lt Mul Rep po Val Sub Sub -0.0062526521217403 Gt Mul po Val 9.320244570715 pl Val Sub Val 0 Lt Mul Sub 0.19505921812091 Lt Mul Hours ph Val Days po Val Sub -0.35232823715884 Lt Mul 0.48087209546602 Lt Mul Mul Days vwap Val 1.6236306970182 Mul ptr Val I I I I I ptr SMA Gt Mul Gt Mul 0.62809943857406 Gt Mul Rep pc Val Hours oh Val I I I I I I I ph Val pl Val Sub Abs Add Gt ph Val I Days pc Val I Days po Val Sub 0 Gt Mul Rep Mul RepN I I I I I I I I I time Val 0.59345924286126 Gt Mul ptr Val Gt Mul I time Delta I po Val pl Val Sub pc Val pl Val Sub Log Min2 0.60342306191225 Gt Mul -0.24898811846294 Gt RepM

VISOR – bitcoin price prediction (Lua, Torch)
In this project I applied deep convolutional networks to predict price movement. I trained different models on a specially constructed bitcoin price chart (images). I managed to train model that gained profit on new data it has never seen before. Though its profit factor is too low to use it for real trading.

Trained on: 01.09.2017 – 01.10.2018.
Tested on: 16.01.2019 – 18.06.2019
Profit factor: 1.28, max drawdown: 8%