LEV SELECTOR, Ph.D.
New York City, USA | Lev.Selector@gmail.com
https://www.linkedin.com/in/levselector/

SUMMARY:

- Artificial Intelligence AI, Large Language Models LLM, Embeddings, Vector Databases, LangChain, Machine Learning, Data Science, Analytics, Expert, Leadership, Team Builder
- Ph.D. in mathematical modeling and computer simulations
- Experience of building teams of engineers and data scientists (hiring, training, and managing)
- 15+ years of experience with high-volume data processing, analytics, and modeling
- Proven track record of building data and analytics systems from scratch, and delivering them on time
- Full cycle development from requirements and planning to architecture & implementation
- Experience in different areas - Financial, Advertising, e-Commerce, Media, Publishing, Crypto
- Data Engineering on clouds (AWS, Azure, GCP) and on-prem, ETL, APIs, big data, data mining
- SQL databases, SQL Data Warehouse, key-value stores (like Redis),
- Web applications, and business intelligence & analytics
- Mathematics and Finance: Ph.D. (Math. Modeling), Advanced Calculus (Math Analysis), Linear Algebra, Vectors & Tensors, Theory of Complex Variables, Differential Equations, Partial Differential Equations, Differential Equations in Mathematical Physics, Probability & Statistics, Statistical Radiophysics (including time series analysis and digital pattern recognition algorithms), Optimization, Monte Carlo Simulations, Numerical Methods, Machine Learning, Deep Learning, NLP, Transformers, Generative AI, LLMs
- Programming & Data Science Tools: Python, Pandas, NumPy, SQL, ETL, Scikit-Learn, NLP/NLTK, Huggingface tools, AWS SageMaker, TensorFlow, Hadoop, Perl, Go (Golang), Javascript, C/C++, Java, Excel VBA, databases (PostgreSQL, MySQL, Vertica, Netezza, Google BigQuery, MongoDB, Sybase, DB2, Oracle, MS SQL), Web Apps, Cloud (Amazon AWS, Azure Synapse Analytics, Google, IBM)

PROFESSIONAL EXPERIENCE:

January 2022 - present - Enterprise AI Solutions, President
- Large Language Models (LLM) for business - local and on cloud
- Custom private conversational interfaces for business systems
- Talk to your data - cognitive search
- Text generation (generate contracts, instructions, summaries, q/a)
- LLM quantizing, fine-tuning. Prompt-engineering. Self-verification
- RAG (Retrieval Augmented Generation), embeddings, vector database, hybrid database, PostgreSQL
- Huggingface tools, models & APIs, llama.cpp, LangChain, OpenAI, ChatGPT, LLaMa, Ollama, LMStudio

November 2022 - Sept 2023 - CTO & Co-Founder, BixBeta
- AI-enabled crypto accounting and payment/donation
- Generative AI to add Natural Language Interface to analytics and reporting

November 2020 - 2022 - Data Science Practice Lead, Redapt
- Building, training, and managing a team of data scientists
- Data architecture in the cloud (Azure, AWS, Google), Apache AirFlow, ADF, SQL Data Warehouse
- Hands on Machine Learning modeling in the cloud (Azure Machine Learning Studio, Azure Synapse Analytics, Azure Cognitive Search, Python, Scikit-Learn, AutoML, predictive modeling, regression, classification, time series forecasting, anomaly detection)

February 2020 - November 2020 - Selectorweb, consulting projects
- National Debt Relief (https://www.nationaldebtrelief.com) Machine Learning and ETL in production
- Provisions Group (https://provisionsgroup.com) Machine Learning with AWS SageMaker and R
- Vara (now https://vareto.com) Data Integration for analytics

December 2019 – February 2020 - Head of Machine Learning And AI, Digital Labs, Capco
- Built, trained and managed a team of data scientists, hands on modeling using Python, AWS, Linux VMs, Mainframe Test Automation project using Python, GnuCOBOL, libFuzzer, antlr4, Neo4j

December 2018 – September 2019 - VP of Data Science, Head of Analytics - National Debt Relief, LLC
- Built, trained and managed a team of data scientists; created architecture on Linux server(s) on Azure cloud to run hundreds of jobs (ETL, data processing, Machine Learning); building visualizations using Salesforce Einstein

Analytics Dashboards and Jupyter notebooks. Azure SQL Data Warehouse, Python, Nginx web server, Gunicorn gateway server, Scikit-learn, Random Forest, XGBoost, H2O.ai)

June 2018 – December 2018 – Senior Data Scientist at Vanguard Group via Tata Consulting Services - Machine Learning and Artificial Intelligence, Scikit-Learn, AWS SageMaker. Designed architecture for data migration into the cloud (AWS) for analytics, developed data structures and lambda functions. Managed deployment. Trained a team in ML and AI (hands-on and lectures). Built models for cybersecurity & anomaly detection.

June 2017 – June 2018 – Selectorweb, consulting projects (Clients: Galvanize, SaleZoom, JKCF) - Machine Learning, AI, Analytics, working with imbalanced data, NLP (Natural Language Processing), Python, Scikit-Learn, NLTK, Pandas, Numpy, TensorFlow, Amazon AWS (EC2, S3), Google Cloud (BigQuery, MySQL), IBM Cloud (IBM Watson – Natural Language Understanding for Sentiment Analysis, Tone Analyzer, Personality Insights).

November 2014 – June 2017 - Penguin Random House, Consultant – Trained and led the team to build systems in Business Intelligence/Analytics, Big Data Collection and Integration (ETL/ELT), Machine Learning and AI, Reporting. Python, pandas, Netezza, SQL, Redis, AWS, Linux. Created system for high-volume parallel web scraping at ~2 Mln pages/day. Organized "Deep Learning Book Club" to promote the use of Machine Learning and AI in-house; worked with several business groups to create various data feeds and tools to clean and ingest data; created Python framework to work with IBM Netezza DB; wrote analytics tools for price analysis and estimation.

April 2012 - April 2014 – AppNexus, Inc, Consultant, Financial Data Analytics - Designed and implemented a billing and reporting framework for processing trading data (~60TB/day) across 2000 clients with custom rules (43 thousand lines of python code, self-recoverable ETL processes, health-monitoring, auto-back-testing to validate both data and code). Designed business intelligence & analytics systems (multiple reports, graphs, analytics database - MySQL). Linux, Python, Pandas, Vertica, Mysql, Hadoop, Hive, Git. Reduced data extraction time for 30 Bln data rows from 3 hours to 7 minutes. Trained and managed a team of developers. Taught courses on python/pandas/database/analytics.

1994-2012 - multiple "Wall Street" consulting projects:

- **Goldman Sachs**
- **Morgan Stanley**
- **JPMorgan Chase**
- **Citigroup**
- **Merrill Lynch**
- **Waterhouse Securities**
- **CantorFitzgerald/Espeed**
- **CSFB**
- **HSBC**
- **WorldQuant**

Data Processing and Analytics using Unix, Perl, SQL (Sybase, DB2, MySQL),
data preparation, data load/migration, ETL,
Web design (HTML, CSS, Javascript), C/C++, Java, Jython, Excel VBA

1991-1994 Columbia University. Staff Associate - Mathematical modeling of dynamics of organic molecules. Distributed calculations (unix, C/C++), analytics and graphics using IGOR software on Mac.

1981-1991 National Cardiology Research Center, Moscow, Russia. Researcher - Real time data acquisition and computer processing in neuro-physiological experiments. Pattern recognition, classification, computer simulations of nerve impulse generation and propagation along C-fibers. Partial differential equations, Hodgkin-Huxley model, Crank-Nicolson & modified Runge–Kutta methods. Hardware and software design of medical equipment

EDUCATION:

- 1988 Ph.D. in Mathematics (modeling of nervous coding), Moscow Institute of Physics and Technology
- 1981 MS in Automation, Moscow Institute of Physics and Technology (MIPT), majoring in computers, electronics and biophysics, Diploma - computer simulation of nerve activity

LICENSES, CERTIFICATIONS, COURSES:

Coursera courses (Machine Learning, Deep Learning, Generative AI, Google Cloud). Data Analysis with Python and Pandas. SEC Registered Representative (Series 7, Series 63). CQF (Certificate in Quantitative Finance). Advanced Object Oriented Perl. C++ for Quantitative Finance. Advanced Excel for Financial Applications. Java 2 (Sun)

REFERENCES:

.. http://levselector.com/LevSelector_recommendations.docx