|  |  |
| --- | --- |
|  |  |

**Candidate 345AB**

Skills

**Academic Disciplines**

* **Social Sciences:** Economics
* **Humanities:** Russian, English, Belarusian
* **Formal Sciences:** Regression Analysis, Prediction Interval, Nonlinear programming, Time Series Analysis (Holt's Exponential smoothing), Statistical Analysis and Modeling, Confidence Interval, ANOVA
* **Applied Sciences:** Acceptance Test-Driven Development, Exploratory Data Analysis, Test-Driven Development, Functional Programming, Neural networks, Structured Logging, Statistics, Quantitative Methods, Virtualization

**Business Functions**

* **Learning and Development:** Application of learning technology, Learning content development, Evaluating learners progress, Learning Programs Management, Manage and Improve L&D Processes, Context analysis and needs assessment, Translating science and learning research into meaningful evidence-based practice and programs, Designing learning solutions, Facilitating group learning process
* **Sales and Marketing:** Text writing
* **Talent Acquisition:** Interviewing candidates
* **Office and Facilities Management:** Literacy, Phone conversation etiquette, Office equipment usage
* **Accounting and Financial:** Financial planning, GAAP accounting, Management accounting, Financial analysis
* **Account Management:** Upsell and Cross-Sell Assistance

**Consulting Practice**

* **Business Consulting:** Enterprise Education & Learning

**Engineering Practices**

* **Digital Engagement:** Microsoft SharePoint
* **Advanced Technology:** UI Prototyping, BA Stakeholders management, Pre-Sale proposal development, AT/Agile, Requirements approval and prioritization awareness, Elicitation and Collaboration, Business analysis information management, Discovery conducting, Solution definition, Business Model(ing), Requirements traceability management, Business analysis activities planning, Business Analysis, Writing Technical Documentation [English], Business Analysis Approach Definition, Requirements Analysis, Business need definition, BA artifacts documentation, C Programming, Understanding the organization, Usage of screen shots in technical documentation, Roadmap and Backlog creation
* **Intelligent Enterprise:** Database Knowledge and SQL Proficiency, Predictive Analysis, Regression Metrics for Machine Learning, Data Science, Machine Learning Model Development, Data Discovery, In-Memory Data Processing, BI Analysis, Natural Language Processing, Built-In Functions, Relations and Modeling, Dimensional Modeling, Data Quality, Numeric Variables, Out-of-Core and Distributed Data Processing, MS SQL. Working with Metadata, Classification Metrics for Machine Learning, Categorical Variables, Data Integration, Data Visualization Essentials, Data Management, Advanced Analytics and Data Science Fundamentals, BI Analytics and Visualization, Data Governance, Reference and Master Data Management, Data Warehousing Fundamentals, Data Profiling, Data Storage, Data Collection and Analysis
* **Cloud:** DevOps, Cloud
* **Quality Engineering:** Component / Integration Testing, Behaviour-driven Testing, Unit Testing, Functional Testing

**Industries**

* **Energy & Resources:** Petroleum Transport
* **Consumer:** Technical Analysis [Travel], Corporate Action Data [Travel], Economic Data Series [Travel], Fundamental Analysis and Ratio Calculation [Travel]
* **Technology, Media & Telecoms:** Electronics, Analytics Data Interpretation
* **Financial Services:** Order execution management, Profit and loss calculation (P&L), Asset allocation

**Leadership & Soft Skills**

* **Leadership:** Professional Development Planning
* **Communication:** Consensus building, Diplomacy, Meeting Facilitation, Assertiveness, Presenting, Conflict management, Negotiations, Business Correspondence, Visual Representation of Information
* **Consultancy:** Consultancy, Data Analysis, Client Relationship Management, Information Analysis, Logical Modelling
* **Business Acumen:** Market orientation, Financial Awareness
* **Teamwork and Collaboration:** Team Management
* **Growth Mindset:** Adaptability
* **Ownership:** Problem-solving, Decision-making

**Managerial**

* **People Management:** Motivation, Navigating EPAM people management processes
* **Project Management:** Collaborating with Resource Management Organization, Customer Communications, Project Management
* **General Management:** Estimation, Quality Management, Brainstorming, Data Modeling, Prototyping, Data Dictionary and Glossary, Document Analysis, Business Model Canvas, Organization Modeling, General Management, Kanban, User stories, Process Modeling, Scrum, Use Cases and Scenarios, Item Tracking, Risk Analysis and Management, Agile, Business Rules Analysis
* **Service Management:** Information Technology Infrastructure Library

**Technologies**

* **Standard:** UML, .xlsx, Azure Virtual Machines, Regular Expressions, Business Process Model and Notation (BPMN), YAML, Git
* **IDE:** IntelliJ IDEA, Visual Studio, PL/SQL Developer, Jupyter Notebook, JetBrains PyCharm, Eclipse
* **Library:** Matplotlib, pip
* **Hardware:** Azure Internet of Things
* **Computer Language:** R Language, Pascal, Java, Visual Basic, Visual Basic for Applications, UNIX shell scripting, FORTRAN, Bash, JavaScript, SQL, C++, IronPython, Assembler [x86], T-SQL, Python
* **Operating System:** Microsoft Windows, Linux, Unix
* **Framework:** Apache Hadoop, Apache Maven, Behave, Apache Spark, Pandas
* **Data:** Power BI Dataflow, PostgreSQL, Analytics and Reporting Tools, R Data Science Ecosystem, Apache Hadoop MapReduce, Oracle RDBMS, Azure Blob Storage, SQL Management Studio, Microsoft SQL Server, Apache Hadoop HDFS, Python Data Science Ecosystem, Data Integration (ETL/ELT) Platforms, RDFS, Amazon Redshift, Enterprise Analytics & Reporting Platforms, Apache Kafka, Protege
* **Other:** Octave, Message Passing Interface, Azure HDInsight, Amazon Athena, CAD/CAM, OpenMP, GIS development Concepts, Microsoft Power Pivot, Instance Configuration Service, Power BI Report Development, Parser generator, Mathematica Wolfram Language
* **Solution:** VIM, Grafana, Mathematica, Graphviz, Agile Lifecycle Management, Gitlab, Oracle VirtualBox, EPAM Project Management Center, Microsoft Power BI, Amazon EC2, Microsoft Access, Enterprise Architect, Jenkins, Confluence, Microsoft Visio, Microsoft Dynamics CRM, WinSCP, Microsoft Outlook, Amazon S3, Oracle SQL Developer Data Modeler, Elasticsearch, Oracle SQL Developer, Amazon EMR, drawio, Jira, TortoiseSVN, Adobe Photoshop, Apache Hive, MS Project, Apache Zookeeper, Adobe Acrobat Professional, Apache Subversion, OpenOffice, GitHub, MS PowerPoint, BitBucket, Visual Studio Codespaces, Microsoft Excel, Selenium, PuTTY, Microsoft Word, Oracle VM
* **Platform:** Amazon Web Services, Microsoft Power View, Office 365, Docker, MATLAB, dbt

Work experience

**Nov-2020 - Till now (Oct-2023)** - Data Analyst, Systems,

**Project Description:** While EPM-DMTM is an umbrella project to host DM capabilities globally, the Data Science stream within it (AI/Skill group of Alexey Yakunin) is focusing on bringing the Data Science methods to bear on possible automation of tasks around workforce management, especially skill modeling, increasing the usability of the existing algorithms, researching the emerging use cases

The Word2Vec-style Semantic Skill Distance (DistSki) algorithm, conceived and developed by Uladzislau Kisin, established the ML/NLP beach-head and the basis for the whole set of potential use cases. It is one of the efforts aimed at making the Skill Graph and various skill-dimensioned activities more usable. The most frequent type of tasks is establishing various skill-facts between various EPAM objects (Position, Project, Employee, Applicant, Learning profile, Video, Presale opportunity etc.) and EPAM skills, thus providing a strong conformed “Skill dimension” enabling cross-system analytics and a number of corporate processes

**Team Size:** 5 Data Scientists, 3 ML Engineers, 2 DevOps

**Project Roles:** Key Developer, Project Coordinator

**Responsibilities:**

* \* performed research of the subject area of skills/competencies modeling, skill supply and demand – as a result (a) the suggested compound indicator with convex aggregation was included in corporate report on Demand-Supply-mismatch; (b) proposed notion of opportunity cost from unstaffed projects also became a component of reporting
* \* Identified multiple external sources of skill information, including structured (ontologies) and unstructured (Bodies of Knowledge, dedicated texts)
* \* performed investigation and EDA of the Skill Dimension (EPAM Skillo Graph) and Skill Facts (EPAM SkillLink) from the EPAM Data Lake, reported and helped alleviate significant consistency issues, suggested and implemented Analytical Views on the main sources, with radical transformation to using skill name as unique identifier, and converting attributes to uniform skill facts suitable for effective cross-domain analytics
* \* implemented the set of NLP-oriented API end-points under the moniker Skillo-DQ (Data Quality), starting from the data quality services for Skills and Skillo, automated translation from/to multiple languages, and the enhancing wrapper with incremental lexical standardization pipeline around the ‘Skillo Enrichment’ end-point, designed to help in placing new Skill candidate terms into Skill Graph. Resulting data product is used by internal systems: SkillHunter, Project, Reporting, Skillo, etc
* \* participated in the corporate-wide initiative on Key Skills establishment; authored an algorithm of personalized Key Skill suggestions, developed transformation pipeline, and template of RM-oriented report on subordinate key skills with extended explain-ability on suggestions, subject to productization; Performing the data-driven hypothesis testing of impact that well-developed key skill profile may have on EPAM processes and average worker bench/productivity balance
* \* developed multiple niche tasks related to data quality statistics for decision-making around skills, as well as Skill-related analytics
* \* developed common tools and analytical datasets in use in the project stream: personal, organizational, and geographical and skill hierarchies in refactored to one-row analytics ready datasets
* \* provided robust consulting to team members on specific data model objects and best practices of query building in existing corporate context
* \* advocated for better deployment observability and co-led DevOps effort to improve it

**Tools and Technologies:** PostgreSQL, Microsoft VS Code, Anaconda, Pandas, spaCy, gensim, scikit-learn, Google Cloud Platform, NLTK, RegEx, Plotly, numpy, scipy, Jupyter Notebook, Git, Confluence, JIRA, ODAHU, DLab, Python, SQL, FastAPI, Tkinter, Selenium, Linux, Kubernetes, CI/CD, Gitlab, Argo

**Sep-2019 - Oct-2020** - Data Analyst, Project Coordinator

**Customer Description:** Retail & Distribution. Consumer Packaged goods in 10 large categories (like Diapers, Feminine Care, Mens Shaving, Laundry, etc.) across more than 100 countries

**Project Description:** build a quick (6 weeks) PoV (prototype) for SOI modeling for their Samanta application based on the identified requirements/user stories provided by P&G stakeholders

Modeling of the consumer behavior as change in volume/quantity consumed in response to changes in prices, distribution and promotion activities (sometimes referred to as Marketing Mix Modeling, MMM, and called Share-of-Input-Modeling, SOIM at P&G) is a Econometric / Data Science capability at the core of the business value of the project

Lead Data Scientist, who won in the initial contest for the right to implement PoC successfully identified the probabilistic Bayesian Multi-level Econometric Regression modeling as a key instrument to successful implementation. This is a demand system estimation problem, using the weekly syndicated Nielsen data panels

Customer, from early stage, also wanted to have a standalone tool to make the results of modeling operational for the P&G users. SOIM Tool is being developed as a web Application (python modeling and back end, JavaScript / React front-end), with two core functionalities: (1) coefficients of regression model, interpretable as elasticities of demand, are made available (view in-tool and export to Excel) to support the identification of Key Business Drivers (KBDs) of the P&G brand managers and other Marketing executives; (2) Scenario-based simulation (what-if analysis) of promotional campaigns allows to review a significant number of possible plans and scenarios vis-à-vis competitive actions, and choose the most suitable one(s) without spending actual resources and time on in-field experiments

Part of the Data Scientists task is to clarify that results of modeling are not unequivocally valid due to the signal-to-noise ratio limitations of the data, as well as multitude of latent variables not directly observable in the data, as well as the potential failure to account for the alternative actors behavior affecting the system of equations

Effort is made to speed up the training of initial models to provide the quicker ramp-up of new internal customers of the product/tool. Intergation is in progress with the P&G-"sister" project SAMANTA, aimed at providing deeper analytics and suggestions to users of that system. Fundamental re-work of the architecture to facilitate scale is undertaken and now being productionalized

**Team Size:** 3 Data Scientists

2 Back end engineers

1 Front end engineer

1 UI Desgner (first 4 months)

1 DM (part time)

**Project Roles:** Data Scientist, Data Analyst, Business, Analyst, Architect, Project Coordinator

**Responsibilities:**

* \* Performed initial research in feasibility of Causal Modeling techniques application to data of the customer
* \* Participated in Arhictectual decisions in several re-designs
* \* Created Architectural diagrams
* \* Created data model
* \* Designed initial API architecture
* \* Developed RPC-style APIs between front- and back-end
* \* Created a stable immutable-folder-based method of model run deployments and results persisting
* \* Applied developer probabilitic regression models to alternative domeains and datasets
* \* Created Jira workflows, including backlog in Kanban, and managed Jira for the first half-year
* \* Performed customer PO communication management
* \* Performed project coordination functions

**Tools and Technologies:** PostgreSQL, Jupyter Notebook, Microsoft VS Code, Git, Jira, Teams, Confluence, Visio, ProjectLibre, MS Office, Python, PyMC3, Bayesian multi-level regression modeling, JavaScript React, FastAPI, SQLAlchemy, xarray, NCDF file format. Coming: TensorFow Probability (TFP)

Education

**Name of the Education Establishment:** UNIVERSITY OF ILLINOIS

**Faculty/College:** College Of Business Administration

**Department:** Economics

**Specialty:** Economics and Finance

**Degree (diploma):** Master

**Name of the Education Establishment:** UNIVERSITY OF COLORADO BOULDER

**Faculty/College:** Graduate School

**Department:** Department of Economics

**Specialty:** INTERNATIONAL AND PUBLIC ECONOMCIS

**Name of the Education Establishment:** BELARUSIAN STATE UNIVERSITY

**Faculty/College:** Radiophysics and Computer Technologies

**Department:** Electronic Mathematical Machines

**Specialty:** Computer Science and RadioPhysics (5 year engineering degree)

**Degree (diploma):** Specialist

**Name of the Education Establishment:** UNIVERSITY OF COLORADO BOULDER

**Faculty/College:** Graduate School

**Department:** Department of Economics

**Specialty:** INTERNATIONAL AND PUBLIC ECONOMICS

**Degree (diploma):** Master

Training Courses

**Discipline:** Machine Learning (Stanford MOOC through Coursera, Andrew Ng), Certificate in College Teaching