# Lukasz Korycki | Resume

219 East Broad Street, Apt. 1108, Richmond, VA, 23219, USA

☎ +18049390044 • ✉ korycki.lukas@gmail.com • 🌐 lukaszkorycki.com

linkedin.com, researchgate.net, scholar.google.com, github.com

## Personal Profile

A motivated PhD student looking for a ML/AI internship. Currently focused on advancing online learning algorithms that are able to effectively aggregate stable patterns and adapt to changes in dynamic environments. Devoted to developing universal approaches and supporting their applications to real-world issues.

At the same time, a software and data engineer with relevant experience from the big data industry, aware of practical limitations and everyday challenges that come with complex data-driven commercial projects.

## Experience

**January 2018 - Now**

**Research Assistant**, *Virginia Commonwealth University*, Machine Learning and Stream Mining Lab, USA.
- Working on algorithms for lifelong/continual learning from stationary and non-stationary data (streaming, class-incremental, task-incremental).
- Tackling catastrophic forgetting in neural networks: hybridizing CNN with random forest, designing adequate loss functions for imbalanced distributions, utilizing lifelong GAN, improving sampling for experience replay, employing intrinsic ensembles.
- Enhancing adaptation to dynamic concepts: proposing adaptive semi-supervised methods based on self-labeling and active learning, applying dynamic diversity in drifting environments, creating committees of change detectors.
- Developing applications of the algorithms (medical imaging).

**September 2015 - December 2017**

**Software Developer**, *Nokia*, Big Data & Network Engineering department, Poland.
- Designing, benchmarking and maintaining algorithms for querying distributed data sources configured on local servers or in the cloud – utilizing the MapReduce paradigm for this purpose, using MongoDB (JavaScript) and Spark (Scala) queries.
- Working with massive amounts of network data stored in heterogeneous databases including standard relational PostgreSQL, column-based HBase and NoSQL MongoDB.
- Developing a highly parallel system based on microservices written in various technologies (Python, Scala).

## Education & Awards

**2018 - Now**

**PhD, Computer Science**, *Virginia Commonwealth University*, USA.
- Research: *Lifelong learning from stationary and non-stationary data*.
- GPA: 4.00/4.00.

**2015 - 2017**

**MSc, Computer Science**, *Wrocław University of Science and Technology*, Poland.
- Thesis: *Semi-supervised methods for evolving data streams classification*.
- GPA: 3.99/4.00.
- Several chancellor's grants for best students.

**2012 - 2015**

**BSc, Computer Science**, *Wrocław University of Science and Technology*, Poland.
- Thesis: *Real-time emotion recognition using facial expressions*.
- GPA: 3.94/4.00.
- Awards: Top 10 Undergraduate Students of the University, The Best Undergraduate Student of the Department of Electronics.

**2010, 2011** Twice honored with the Prime Minister's grant for best high-school students.

## Selected Publications

**2020** **Online Oversampling for Sparsely Labeled Imbalanced and Non-Stationary Data Streams**.
Oversampling for non-stationary environments with dynamic concepts and class ratios.
*International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020*.

| 2019 | **Active Learning with Abstaining Classifiers for Imbalanced Drifting Data Streams**.<br>A novel query strategy focusing on requesting labels for minority instances.<br>*IEEE International Conference on Big Data, Los Angeles, CA, USA, 2019.* |
|---|---|
| 2019 | **Unsupervised Drift Detector Ensembles for Data Stream Mining**.<br>An improvement over simplistic univariate unsupervised statistical tests for drift detection.<br>*IEEE International Conference on Data Science and Advanced Analytics DSAA, Washington D.C., USA, 2019.* |
| 2018 | **Clustering-Driven and Dynamically Diversified Ensemble for Data Streams**.<br>Applying dynamic diversity to enhance adaptation to changes in streams.<br>*IEEE International Conference on Big Data, Seattle, WA, USA, 2018.* |
| 2017 | **Combining Active Learning and Self-Labeling for Data Stream Mining**.<br>Utilizing unlabeled instances while learning under limited supervision.<br>*International Conference on Computer Recognition Systems, Wrocław, Poland, 2017.* |

## Preprints

| 2020 | **Adversarial Concept Drift Detection under Poisoning Attacks for Robust Data Stream Mining**.<br>The first drift detector distinguishing between normal and adversarial changes.<br>*arXiv:2009.09497 [cs.LG]* |
|---|---|
| 2020 | **Instance Exploitation for Learning Temporary Concepts from Sparsely Labeled Drifting Data Streams**.<br>Strategies exploiting labeled instances to prevent underfitting in sparsely labeled streams.<br>*arXiv:2009.09382 [cs.LG]* |

## Projects

| 2020 - Now | **ML++**.<br>Developing a framework for lifelong/continual machine learning (PyTorch, TensorBoard). The goal is to popularize the generic paradigm of learning (online over offline) by simplifying access to methods that prevent catastrophic forgetting and improve adaptation to non-stationary data. Besides, the framework will provide a unified benchmarking environment for standardized evaluation of novel lifelong learning algorithms. |
|---|---|
| 2018 - Now | **Machine learning algorithms**.<br>Implementations of state-of-the-art classifiers, created to gain deeper insights into them. Including: neural networks with backpropagation (Python), streaming decision rules for Flink (Scala), distributed k-NN for Spark (Scala). |
| 2016 | **OGR (Optical Graph Recognition)**.<br>A system for recognizing graphs manually drawn on a blackboard or piece of paper. The core part of the application is a module responsible for a visual analysis of a graph – finding nodes, connections and reading text (OpenCV, Tesseract). The project was created in collaboration with Nokia for The Student Projects Conference 2015. |

## Key Skills

| ML & AI | All state-of-the-art classifiers (neural networks/deep learning, decision trees, naive Bayes, k-NN), ensembles (bagging, boosting, random forest), semi-supervised and unsupervised learning, incremental learning, algorithms for computer vision and NLP. |
|---|---|
| Programming | Python, Java, Scala, C/C++, JavaScript, object-oriented and functional programming, algorithms and data structures, parallelization, TDD/BDD approaches. |
| Tools/Frameworks | PyTorch, Keras, scikit-learn, numpy, Spark, Flink, OpenCV, basics of CUDA, Git. |
| Databases | PostgreSQL, MySQL, MongoDB, HBase. |
| Mathematics | Solid foundation in: statistics, linear algebra, mathematical analysis, combinatorics. |
| Soft skills | Experienced in working with teams, flexible, responsible, innovative. |
| Languages | English (fluent), German (basic), Polish (native). |