Luca Bazzetto

Data Science Student

+39 3487898075 | luca.bazzetto@gmail.com | linkedin.com/in/lucabazzetto | lucabazzetto.github.io

Education

University of Milano-Bicocca — Milan, IT
MSc in Data Science — *02/2026*

Cracow University of Technology — Krakow, PL
Erasmus+ Exchange Program, Computer Science — *10/2024 – 06/2025*

University of Udine — Udine, IT
BSc in Computer Science — *10/2022*

Experience

Junior Business Intelligence Analyst (Internship) — Treviso, IT
Sinesy srl — *04/2022 – 08/2022*

- Developed predictive models for retail sales forecasting using ARIMA models
- Preprocessed and prepared large datasets, handling missing values, outliers, and seasonality with Python
- Automated data cleaning and transformation workflows, creating reusable scripts for sales data from multiple regions and stores
- Designed a scalable pipeline to generate weekly forecasts for different retail categories and locations, which was integrated into Qlik platform
- Conducted statistical analysis to evaluate model performance

Projects

Text Classification |*Python, TensorFlow* [Repository]

- Developed a deep learning system to classify scientific articles from arXiv using CNN and LSTM architectures
- Implemented K-Fold Cross Validation to evaluate model performance across different architectures
- Created data retrieval and preprocessing pipelines for scientific text data
- Stored and analyzed results from different model architectures to determine optimal approaches

Movie Data Scraper | *Python, BeautifulSoup4, Selenium, MongoDB* [Repository]

- Developed a web scraping tool to collect movie data including ratings and votes from multiple sources (IMDb, MyMovies, ComingSoon)
- Implemented data storage in MongoDB for efficient retrieval and analysis of movie information
- Created modular codebase with separate scripts for different data sources, enhancing maintainability
- Designed command-line interface allowing flexible data collection by country and season parameters

Smoker Detection | *KNIME Analytics Platform* [App]

- Designed data preprocessing workflows including handling missing values, feature engineering, and collinearity reduction
- Trained and optimized multiple models (Random Forest, Gradient Boosting Tree, Logistic Regression, Naive Bayes) using K-fold cross-validation
- Deployed strategies to prevent data leakage, optimize feature aggregation, and address class imbalance
- Implemented manual and automatic deployment strategies for seamless integration of the final model

Awards

KNIME Machine Learning Challenge - 2nd Place [Badge] — Milan, IT
05/2024

Technical Skills

Programming Languages: R, Python, Java, C++, SQL, JavaScript, HTML/CSS, PHP
Developer Tools: Git, VS Code, PyCharm, Tableau, Qlik
Libraries: pandas, NumPy, Matplotlib, TensorFlow, Scikit-learn
Languages: English, Italian