# Data Science Toolkit and Applications

*(Laurent Brisson et Lina Fahed)*

This course covers the entire data science pipeline, from initial data exploration to the implementation and evaluation of machine learning models. It emphasizes hands-on experience with essential tools and libraries, preparing students to tackle real-world data science challenges. The course also incorporates best practices in software development and version control.

Key concepts covered:

- Python programming for data science (including key libraries like Pandas, NumPy, Scikit-learn)
- Data manipulation and storage (structured data formats, databases)
- Data exploration and preprocessing techniques
- Data visualization principles and tools
- Machine learning fundamentals and model evaluation
- Software development best practices (version control, testing, documentation)

By the end of this course, students will be able to:

- [BC-01] Apply best practices in software development to data science workflows
- [BC-02] Effectively communicate insights derived from data through visualization and statistical analysis
- [BC-07] Design and implement end-to-end data science projects using industry-standard tools and libraries

Prerequisites :

- Being familiar with Python programming
- Being familiar with basic Linux commands
- Having basic skills in statistical analysis