WAB

Provadis School of International Management and Technology

Exposé

Examining Random Forest vs Neural Networks for LogP Prediction of Drug-like Compounds

A Proof-of-Concept Implementation and Performance Analysis

Rubin Chempananickal James
rubin.chempananickal-james@stud-provadis-hochschule.de

Department: Information Technology
Module: Fortgeschrittene Programmierung
Reviewer: Prof. Dr. Henrik Paul

February 17, 2026

Contents

Exposé

Problem Statement

The accurate prediction of logP is crucial in various fields, including drug discovery and environmental chemistry.

Objectives

The primary objectives of this project are:

- To find a curated dataset of molecules and find whether a fingerprint-based logp estimation is feasible
- To evaluate the performance of the implemented models in terms of accuracy and efficiency.

Methodology

The project will follow these steps:

1. Data Collection: Gather a dataset of chemical compounds with known logP values. The ZINC20 database will be used as a primary source for this data.
2. Fingerprint Generation: The molecules will be converted to fingerprints using multiple algorithms, like Morgan (ECFP4) and MACCS Fingerprints.
3. Model Development: Train multiple machine learning models like Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN)

Planned Structure

The paper will likely be structured as follows:

- Introduction
- Research Question and Objectives

- Literature Review
- Methodology
- Results and Discussion
- Limitations
- Conclusion and Future Work
- References
- AI Declaration
- Declaration of Authorship