

# Abstract

As starting point the definition of Risk as the chances of having an unexpected or negative outcome has been introduced.

After a brief introduction on most of the risk categories as Banks and regulators, the thesis focuses on credit risk models where the entire financial system is highly investing to avoid a further financial crisis.

Among the Credit Risk metrics, Risk Weighed Assets (RWAs) can be considered an important measure in the current credit risk environment.

Indeed they represent an aggregated measure of different risk factors affecting the evaluation of financial products.

The credit risk model accuracy, as all models, does not depend only on the effectiveness, parametrization and complexity of the model, but from the data used as input. This situation is often summarized as "Garbage IN is equal to Garbage OUT".

In the second chapter, several machine learning techniques for data anomalies detection have been introduced with a focus on Local Outlier Factor (LOF) and Isolation Forests.

In the third and fourth chapters, these algorithms have been tested first on an artificial sample in order to show their statistical properties and then they have been applied on a real credit risk dataset where RWAs data anomalies have been analyzed.