Ministry of Higher Education
and Scientific Research

Academic Year
2024/2025

EPI DIGITAL SCHOOL
Private International Higher
School of Polytechnic

End of Year Project Report

Field of study: AI & Data Science Engineering

Entitled

Forecasting Bitcoin Prices with XGBoost, Prophet, K-Means, and LSTM

Internship place
EPI Digital School

Author
Majd Aguir

Supervisor
Dr. Boutheina BEN ISMAIL

Acknowledgements

I would like to express my deepest appreciation to all those who, from near or far, supported this project throughout its development.

First and foremost, I would like to sincerely thank ***Mrs. Boutheina Ben Ismail*** *for her invaluable guidance, encouragement, and insightful contributions. Her expertise and support played a crucial role in successfully realizing this work.*

I also extend my heartfelt thanks to my family and friends, whose unwavering support and encouragement have been a constant source of strength throughout my academic journey. Their belief in my abilities has continuously inspired me to persevere and strive for excellence.

Finally, I would like to express my most sincere gratitude to the members of the jury for having honored me by agreeing to evaluate this project. I hope that this work meets their expectations in terms of clarity, rigor, and motivation.

List of Abbreviations

XGBoost	eXtreme Gradient Boosting
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
R^2	R-squared (Coefficient of Determination)
LSTM	Long Short-Term Memory
CRISP-DM	Cross-Industry Standard Process for Data Mining
RNN	Recurrent Neural Network
API	Application Programming Interface
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning

Abstract

This project explores the use of machine learning models to predict Bitcoin prices, addressing the challenges posed by its high volatility and complex behavior. The models applied include XGBoost, Prophet, K-Means clustering, and Long-Short-Term Memory (LSTM).

Historical Bitcoin price data was collected and preprocessed to generate features such as lag values, rolling averages, and temporal indicators. The data set was divided into training and test sets, and each model was optimized and evaluated using mean absolute error (MAE), root mean square error (RMSE), and R^2.

The results show that XGBoost captured complex patterns effectively, Prophet modeled seasonality well, K-Means identified price trends, and LSTM excelled at learning sequential dependencies. Combining these models may lead to more accurate and robust predictions.

This research highlights the potential of machine learning in financial forecasting, offering valuable tools for investors and analysts in the cryptocurrency market.

keywords: Machine learning, Bitcoin, financial forecasting, XGBoost, Prophet, K-Means, LSTM.

Résumé

Ce projet explore l'utilisation de modèles d'apprentissage automatique pour prédire les prix du Bitcoin, en réponse à sa forte volatilité et à la complexité de son comportement. Les modèles utilisés incluent XGBoost, Prophet, le clustering K-Means et les réseaux de neurones LSTM.

Les données historiques du Bitcoin ont été collectées et prétraitées pour générer des caractéristiques telles que les valeurs de décalage, les moyennes mobiles et les variables temporelles. Le jeu de données a été divisé en ensembles d'entraînement et de test, et chaque modèle a été optimisé et évalué à l'aide de l'erreur absolue moyenne (MAE), de la racine de l'erreur quadratique moyenne (RMSE) et du coefficient de détermination (R^2).

Les résultats montrent que XGBoost capture efficacement les schémas complexes, Prophet modélise bien la saisonnalité, K-Means identifie les tendances historiques, et LSTM apprend les dépendances temporelles. La combinaison de ces modèles pourrait améliorer la précision des prévisions.

Cette recherche met en évidence le potentiel de l'apprentissage automatique pour la prévision financière, en offrant des outils utiles aux investisseurs et analystes du marché des cryptomonnaies.

mots-clés: Apprentissage automatique, Bitcoin, prévision financière, XGBoost, Prophet, K-Means, LSTM.

Contents

List of Figures

List of Tables

General Introduction

The rise of digital transformation has fundamentally reshaped the way industries operate, with data-driven decision-making now playing a central role across sectors such as finance, healthcare, and technology. Among the most transformative developments in recent years is the emergence of cryptocurrencies, with Bitcoin standing out as the most prominent and widely adopted. Originally conceived as a decentralized alternative to traditional currencies, Bitcoin has evolved into a significant financial asset, attracting the attention of individual investors, institutional players, and regulatory bodies alike.

Despite its growing importance, Bitcoin remains highly volatile, with price fluctuations driven by a complex interplay of global economic factors, regulatory news, investor sentiment, and technological innovations. This volatility presents a major challenge for financial analysts and investors who seek to forecast future market trends with accuracy and reliability. Traditional statistical models often fall short in capturing the non-linear and dynamic nature of cryptocurrency markets, leading to a growing interest in more advanced forecasting techniques.

In this context, machine learning and deep learning methods offer a powerful alternative. By leveraging large volumes of historical data and identifying subtle patterns within it, these techniques provide new opportunities for predictive analytics in the financial domain. This project investigates the application of four prominent models—XGBoost, Prophet, K-Means clustering, and Long Short-Term Memory (LSTM) networks—to forecast Bitcoin prices. Each of these models brings unique strengths, from modeling temporal trends and seasonality to capturing sequential dependencies and identifying historical price regimes.

The goal of this project is to develop a robust and accurate forecasting system that enhances our understanding of Bitcoin price behavior and supports informed decision-making in the cryptocurrency market. To achieve this, historical price data is collected and preprocessed to create relevant features, and each model is trained and evaluated using standard performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error

(RMSE), and R^2 score.

This report is structured as follows: Chapter One introduces the project and outlines its background, motivations, and objectives. Chapter Two provides a comprehensive review of related work and the theoretical foundations of the selected models. Chapter Three presents the methodology used, including data collection, preprocessing, and model implementation. Chapter Four discusses the results and evaluates the performance of the models. Finally, the report concludes with a general summary and reflections on the findings, including suggestions for future research and improvements.