

B.Sc. COMPUTER SCIENCE WITH INDUSTRIAL PLACEMENT  
COMPUTER SCIENCE DEPARTMENT

## **Machine Learning to Study Patterns in Chess Games**

CANDIDATE

**Isaac Cheng**

Student ID 690065435

SUPERVISOR

**Chico Camargo**

University of Exeter

ACADEMIC YEAR  
2022/2023

## Abstract

Chess, one of the oldest and most popular board games, has recently seen large-scale exploration driven by online platforms like Lichess and Chess.com. This study uses data mining and machine learning techniques to analyse millions of games from the Lichess Open Database to uncover patterns and insights into how people play chess. We created a data pipeline for efficient processing, explored data features and distribution, and performed feature engineering. We implemented classification models to predict game outcomes, a regression model to investigate opening popularity and game outcomes, and k-means clustering to group openings by game outcomes and mean differences in their variations. Our models and clusters were analysed for their usefulness and evaluated for their ability to provide insights into chess game patterns. Through this evaluation, we assessed their success, limitations, and potential improvements. This study demonstrates the potential of data mining and machine learning techniques in uncovering patterns and insights in chess, contributing to growing research. By understanding how people play chess, we can develop better tools, strategies, and educational resources, enhancing fairness and enjoyment for players worldwide.

	Yes	No
I certify that all material in this dissertation which is not my own work has been identified.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
I give the permission to the Department of Computer Science of the University of Exeter to include this manuscript in the institutional repository, exclusively for academic purposes.	<input checked="" type="checkbox"/>	<input type="checkbox"/>