

# Relazione progetto machine learning

Martino Pettinari 866496, Davide Creati

February 5, 2024

## Abstract

Nella relazione abbiamo cercato di riassumere tutti gli step che ci hanno permesso di arrivare all'obiettivo del progetto ovvero addestrare 2 modelli scelti e valutarne la correttezza e precisione. Nel processo si includono scelte stilistiche, assunzioni o ipotesi fatte durante lo svolgimento.

## 1 Introduzione e obiettivi

Per la realizzazione di questo progetto, ci siamo concentrati nel lavorare con dati provenienti dal mondo del tennis. La nostra scelta di concentrarci su questo sport è stata guidata principalmente da una ricerca effettuata. Dall'analisi è emerso che, tra l'infinità degli sport che vengono praticati al giorno d'oggi, il tennis offre ottime possibilità per il mondo del machine learning. Ciò è dato dal fatto che è composto da un ambiente meno complesso, un esempio è dato dal fatto che ogni match viene giocato da soli 2 giocatori che si sfidano (dato che nel nostro caso abbiamo considerato solo match singoli e non in doppio).

Il secondo passaggio, una volta identificato il macro perimetro, è stata identificare dei tornei che ci permettessero di avere dati recenti e aggiornati. Per fare ciò abbiamo deciso di analizzare e seguire i tornei del circuito professionistico ATP per il quale, ogni anno, vengono giocati 4 Grand Slam. Il circuito professionistico ATP è composto da molteplici tornei che si svolgono durante l'arco dell'anno. Ognuno di questi tornei ha punteggi diversi, in base all'importanza del torneo.

Per fini progettuali ci siamo focalizzati solo sui tornei Grand Slam (ovvero quelli con il maggior punteggio) avvenuti nel 2023. Questa decisione è stata guidata dal fatto che siano eventi in cui i dati hanno una maggiore consistenza. Questo poiché i giocatori offrono prestazioni al massimo delle loro capacità dato che, la posta in gioco (rappresentata dai punti assegnati), crea un incentivo per i giocatori, garantendo partite di alta intensità e competizione.

L'obiettivo centrale di questo progetto è stato sviluppare 2 modelli scelti di modelli di machine learning in grado di prevedere l'esito di ogni singola partita disputata durante l'ultimo Grand Slam del 2023, utilizzando i dati degli altri 3 tornei per l'addestramento.

## 2 Dati utilizzati