

Tecniche di text analytics e machine learning per la classificazione automatica di giochi da tavolo basata su manualistica, categorie e meccaniche di gioco

Introduzione al
problema

Tecnologie e
strumenti

Prima fase

Reperimento
manualistica

Seconda fase

Terza fase

Machine
Learning

Quarta fase

Conclusioni e
sviluppi futuri

Ringraziamenti

Università degli studi di Modena e Reggio Emilia
13 aprile 2022

Candidato
Dattolo Simone

Relatore
Prof. Riccardo Martoglia

Introduzione al problema

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

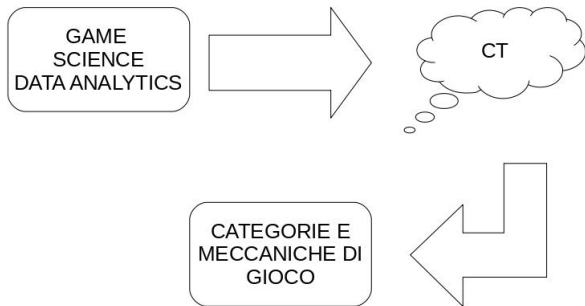
Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

Figura: L'ambito di studi



Introduzione al problema

Introduzione al
problema

Tecnologie e
strumenti

Prima fase

Reperimento
manualistica

Seconda fase

Terza fase

Machine
Learning

Quarta fase

Conclusioni e
sviluppi futuri

Ringraziamenti

- Approfondire la branca legata alla **Game Science Data Analytics**
- Progetto GameInLab, dal *Computational Thinking* e i fattori che lo influenzano
- Direzione verso **le meccaniche e categorie di un boardgame**
- Possibilità di ricavare moltissimi dati a riguardo grazie a **BGG**

Focus sui problemi

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

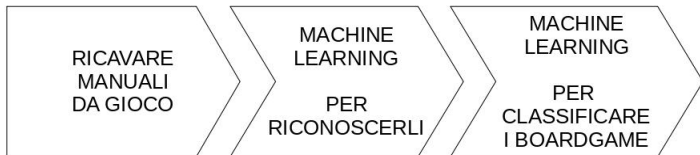
Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

Figura: I problemi da trattare



Tecnologie e strumenti

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

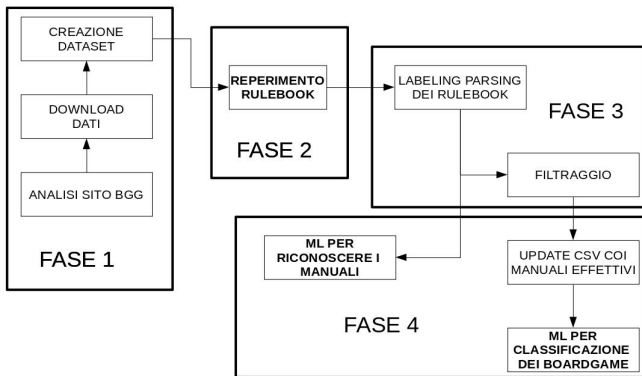
Conclusioni e sviluppi futuri

Ringraziamenti

- Python
- Jupyter
- **Selenium**
 - Automatizza il browser, usato su Firefox
- **Machine Learning**

Panoramica

Figura: Panoramica progettuale



Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

FASE 1 Analisi, download e creazione dati

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

- Uso del ranking BGG
- Ogni boardgame possiede schede consultabili
- Download di ogni boardgame, tramite API
- Parsing dei file associati XML, costruzione di un dataset csv

FASE 2 BGG per scaricare i manuali

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

- Per ogni boardgame nel ranking, è presente una scheda *Files*
- Ci sono file associati, caricati dalla community
- Possibile che ci sia il manuale. Ordinamento dei files per thumbs up e inglese
- **Problema: obbligo di essere loggati sul sito per vedere i link alle risorse**

La scheda files

Figura: Scheda file di un boardgame

Overview Ratings Forums Images Videos **Files** Stats Versions Expansions My Games Market More ▾

Files Hot Recent English ▾ + Upload

Showing 1-25 of 31 ☰ ☰ ☰ ☰ ☰

- ▲ 113 | ● 4 [PDF Official rules of 7 Wonders Duel](#)
Official rulebook of 7 Wonders Duel in English
AntoineReposProd · 6 years ago · English
- ▲ 106 | ● 9 [PDF Brief reference in size of a Wonder card](#)
Two sided small reference. Print 2 sides separately, cut out and put inside of a wonder size card sleeve.
Version 4: added background (thanks indianajo883 for idea)
basta · 6 years ago · English
- ▲ 71 | ● 14 [PDF 7 Wonders Duel Complete Reference Sheet](#)
A complete reference sheet for 7 wonders Duel with expansion and promos.
V4 Grammatical changes
sainttain · 5 years ago · English
- ▲ 59 | ● 10 [PDF Simple, quick solo variant for 7 Wonders Dual](#)
A quick AI solo variant for 7 Wonders Dual.
fattaf23 · 5 years ago · English
- ▲ 50 | ● 5 [PDF 7 Wonders: Duel - Printable Score Sheets](#)
I can personally see myself getting through the supplied pad rather quickly so here are 4 Score sheets per A4 page closely matching the design and dimensions of the original....
PinkyMan · 6 years ago · English
- ▲ 41 | ● 5 [PDF 7 Wonders Duel: A Plain and Simple Guide](#)
Also now added an A5 super summary, shortening the rules to a double sided A5 version.
malbu_bebe_28 · 6 years ago · English

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

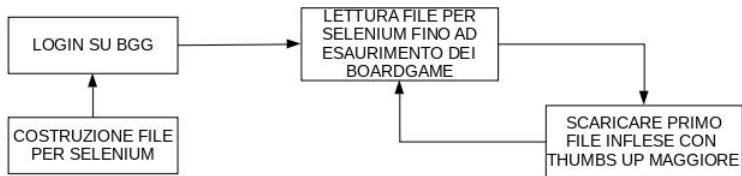
Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

Selenium per il download



Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

FASE 3 Labeling e Filtraggio

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

- Fase di labeling e successivo filtraggio
- Apertura di ogni singolo file scaricato → labeling True/False
- Mantenimento di tutti, per **machine learning per riconoscere i rulebook**
- Aggiunta del campo rulebook al csv precedente per effettuare **machine learning per classificare i boardgame**

FASE 4 Machine Learning: setting esperimenti

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

- Preprocessing
- Resampling
- ML riconoscimento manuali
 - 2.8k campioni, 1.4k per ogni classe.
- ML classificazione
 - mantenimento per ogni gioco della meccanica/categoria più frequente
 - meccaniche → 5 meccaniche più frequenti, 250 campioni per ogni classe
 - categorie → 5 categorie più frequenti, 300 campioni per ogni classe

ML per riconoscimento di manuali

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

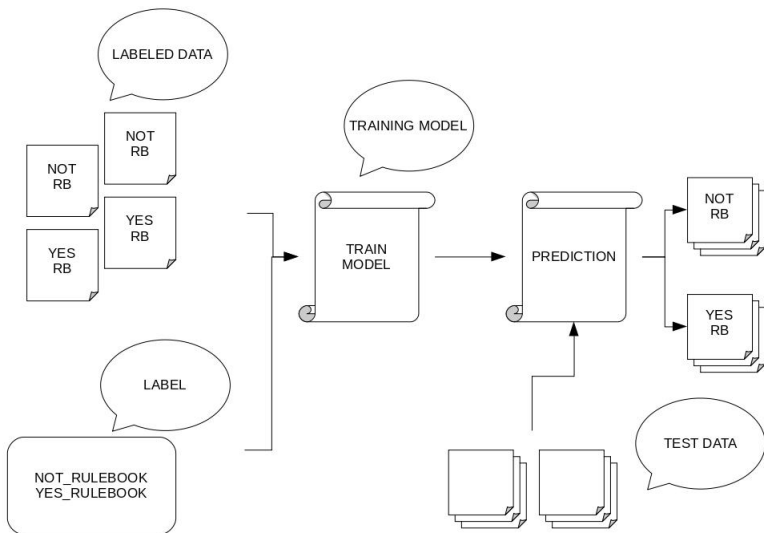
Terza fase

Machine Learning

Quarta fase

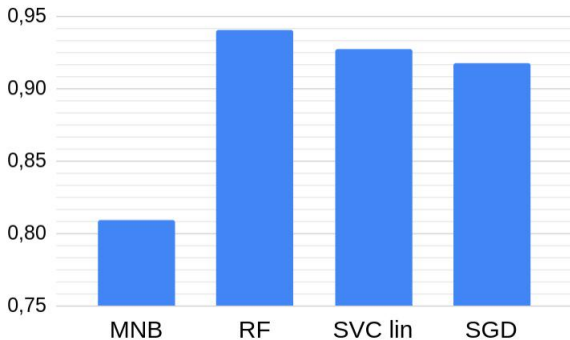
Conclusioni e sviluppi futuri

Ringraziamenti



Risultati

Figura: Accuracy media CV



Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

ML classificazione per meccaniche

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

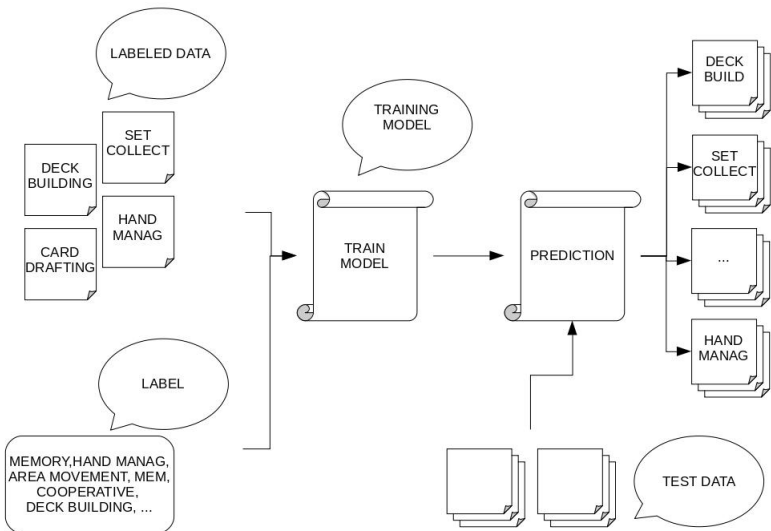
Terza fase

Machine Learning

Quarta fase

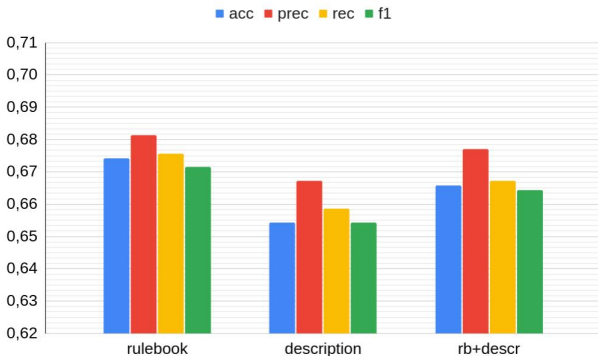
Conclusioni e sviluppi futuri

Ringraziamenti



Risultati

Figura: Media delle medie di tutti gli algoritmi



Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

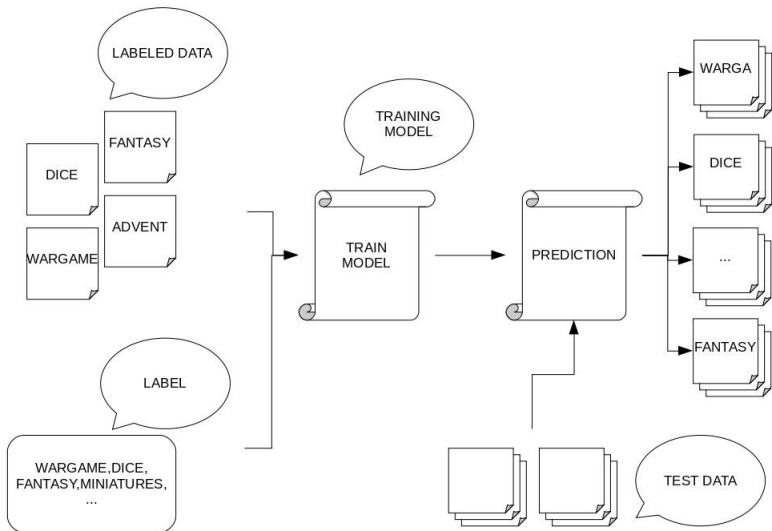
Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

ML classificazione per categorie



Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

Risultati

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

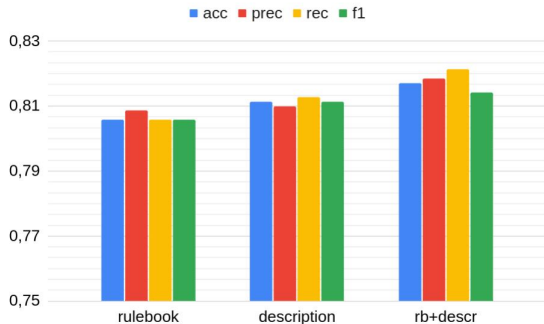
Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

Figura: Media delle medie di tutti gli algoritmi



Most Important Feature

Introduzione al
problema

Tecnologie e
strumenti

Prima fase

Reperimento
manualistica

Seconda fase

Terza fase

Machine
Learning

Quarta fase

Conclusioni e
sviluppi futuri

Ringraziamenti

- Feature interne a un campo di testo, o una variabile, che spiegano come influiscono sul modello
- Focus su metodi ensemble usati → Random Forest

Figura: Most Feature classificazione meccaniche,rulebook

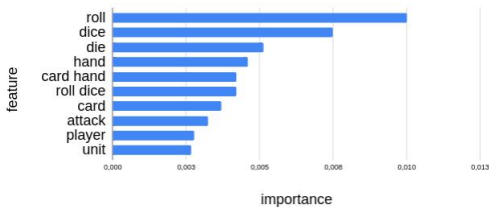
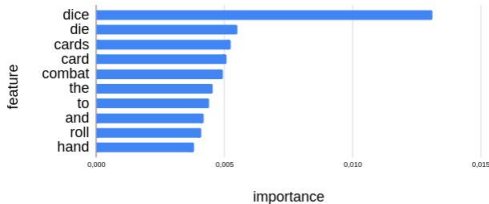


Figura: Most Feature classificazione meccaniche,description



Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

Figura: Most Feature classificazione meccaniche, rb+description

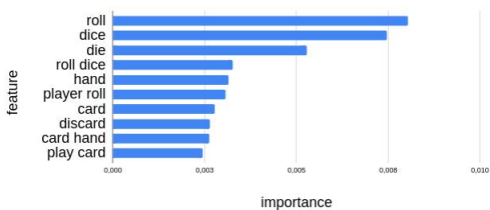
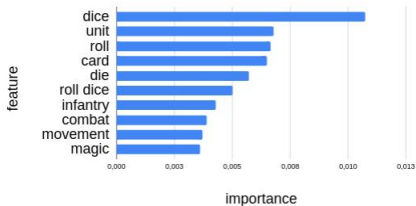


Figura: Most Feature classificazione categorie, rulebook



Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

Figura: Most Feature classificazione categorie, description

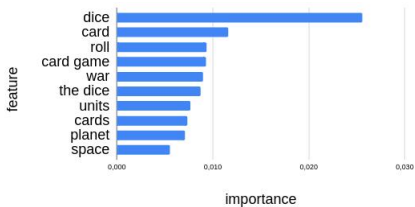
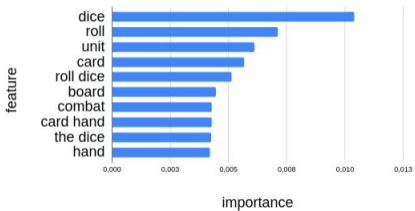


Figura: Most Feature classificazione categorie, rb+description



Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

Conclusioni

Introduzione al problema

Tecnologie e strumenti

Prima fase

Reperimento manualistica

Seconda fase

Terza fase

Machine Learning

Quarta fase

Conclusioni e sviluppi futuri

Ringraziamenti

- Classificazione basata su meccaniche migliore usando il rulebook
- Classificazione basata su categorie migliore usando rulebook e descrizione assieme
- Feature particolari **dice,roll**
- Reti neurali
- Multilabeling

Ringraziamenti

Introduzione al
problema

Tecnologie e
strumenti

Prima fase

Reperimento
manualistica

Seconda fase

Terza fase

Machine
Learning

Quarta fase

Conclusioni e
sviluppi futuri

Ringraziamenti

Grazie per l'attenzione!