

A decorative graphic on the left side of the slide, consisting of a network of thin, light blue lines and small circles, resembling a circuit board or a stylized tree structure.

# INVESTIGATING A STEAM DATASET

STUDENT: CONOR MCCAULEY

SUPERVISOR: PROF. DOUG LEITH

# THE STEAM PLATFORM

- Video game distribution platform
- Largest of its kind on PC
  - 60 thousand games
  - 120 million monthly users
- Games purchased and launched from the desktop app
- Users can review games they've played
  - Binary rating ('thumbs up' or 'thumbs down')
  - Written review

# THE STEAM PLATFORM

- Social network functionality
- Add other users as friends
- 'Community Hubs' for games
- User groups

# STEAM: STORE PAGE

All Games > Action Games > Apex Legends Official Franchise > Apex Legends™

## Apex Legends™

Community Hub



Apex Legends is the award-winning, free-to-play Hero Shooter from Respawn Entertainment. Master an ever-growing roster of legendary characters with powerful abilities, and experience strategic squad play and innovative gameplay in the next evolution of Hero Shooter and Battle Royale.

RECENT REVIEWS: **Very Positive** (24,889)

ALL REVIEWS: **Very Positive** (400,987)

RELEASE DATE: 5 Nov, 2020

DEVELOPER: Respawn Entertainment

PUBLISHER: Electronic Arts

Popular user-defined tags for this product:

[Free to Play](#) [Battle Royale](#) [Multiplayer](#) [Shooter](#) [+](#)

# STEAM: GAME REVIEW



1,336 products in account  
129 reviews



**Recommended**  
270.0 hrs on record (135.4 hrs at review time)

POSTED: 25 FEBRUARY

*Apex Legends is actually an extremely fun game, used to love playing it with friends. Especially in its early days and truly this game has come a long way~ x*

Was this review helpful?

 Yes

 No

 Funny

 Award

33 people found this review helpful  
1 person found this review funny







 2

3 

# THE DATASET

- Scraped from Steam
- Data for 4 million users
- Friends
- Group membership
- Reviews written



# THE DATASET

- Reviews written:
  - Rating (thumbs up/down)
  - Text (multiple languages)
  - Timestamp
  - Playtime ('at review'/'total')
  - Votes (helpful/funny)

# RESEARCH APPROACHES

- I. Relationship between review text and other features
- II. Relationship between a game's early reviews and future success
- III. Influenced behaviour among users and particularly influential users



# APPROACH I

- Relationship between review text and ...
  - Rating
  - Playtime
  - Votes
- Can the review text predict these features?
- Use BERT to vectorise the text for training a deep learning model
- Use as classifier (rating) or regressor (votes)

# APPROACH I

- BERT:
  - State-of-the-art technique in NLP
  - Maps tokens in text to vectors
  - Accounts for context of words
  - Fine-tune pre-trained model(s) on our dataset
  - Use TF-IDF as a baseline vectoriser
  - Use Naive Bayes, SGD as baseline classifiers/regressors

## APPROACH II

- Relationship between early reviews and future success
- How to determine success?
  - Number of players (reviews)
  - Playtime
  - Percentage of positive ratings
- BERT used for text vectorisation

## APPROACH III

- Influenced behaviour among users and particularly influential users
- Determine potentially influenced behaviours:
  - User reviews a game that a friend already reviewed
- Can influenced behaviours be predicted?
- Can we use the wider social graph to find “influencers”?

An abstract graphic on the left side of the slide, consisting of a network of thin, light blue lines and small circles, resembling a circuit board or a neural network diagram. The lines and nodes are concentrated on the left edge and extend slightly into the main content area.

ANY QUESTIONS?