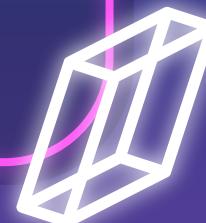


# Predicting Difficulty of Stepmania Charts

With Machine Learning

Svitlana Glibova



# TABLE OF CONTENTS

01

What is Stepmania?

02

Data Collection

03

Modeling / Results

04

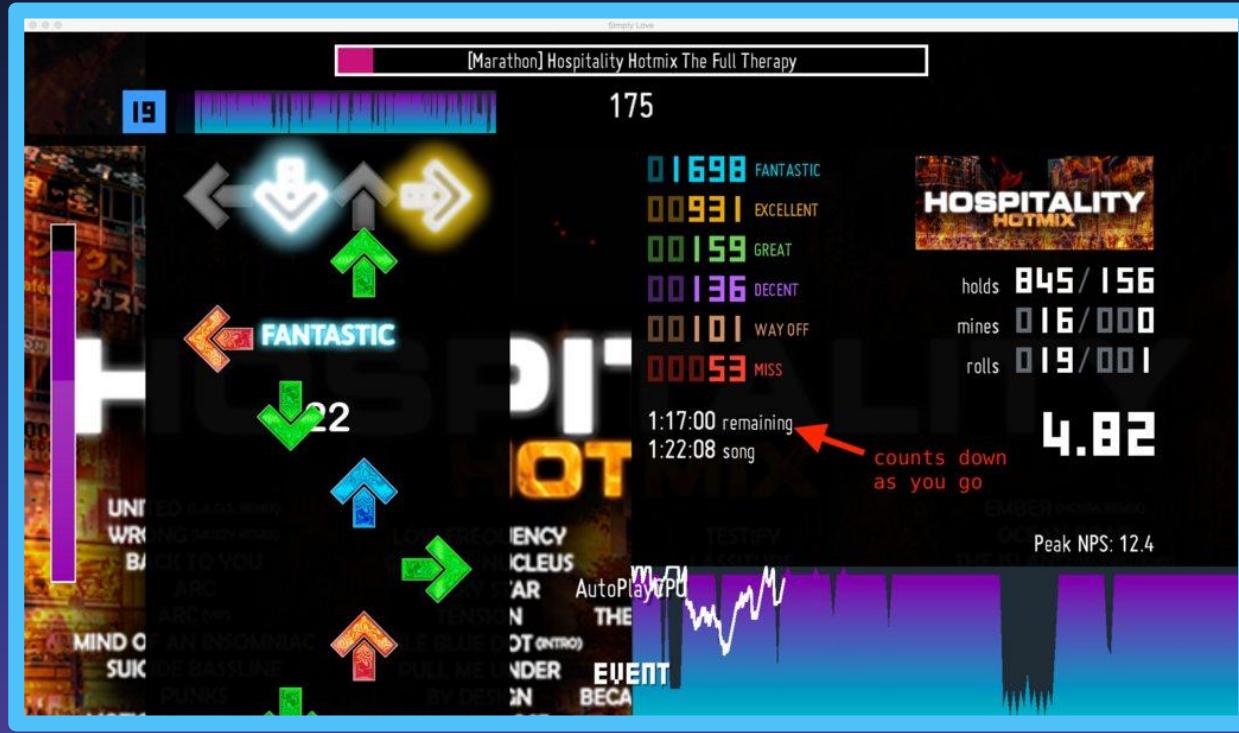
Next Steps

01

# What is Stepmania?

- Open-source software for music games, created in 2005
- Endlessly customizable
- Most popular for “4-panel” dance games (think Dance Dance Revolution)





02

## Data Collection

[ITGPACKS.COM](http://ITGPACKS.COM) - community-driven spreadsheet

Chart parser built by [Tim Murphy](#), fellow  
dance-gamer and software engineer

3200 Technical songs  
2250 Stamina songs

# Technical vs. Stamina Data

## Technical

- Pattern complexity - more technical features
- Shorter songs

Target: Rating - numerical difficulty

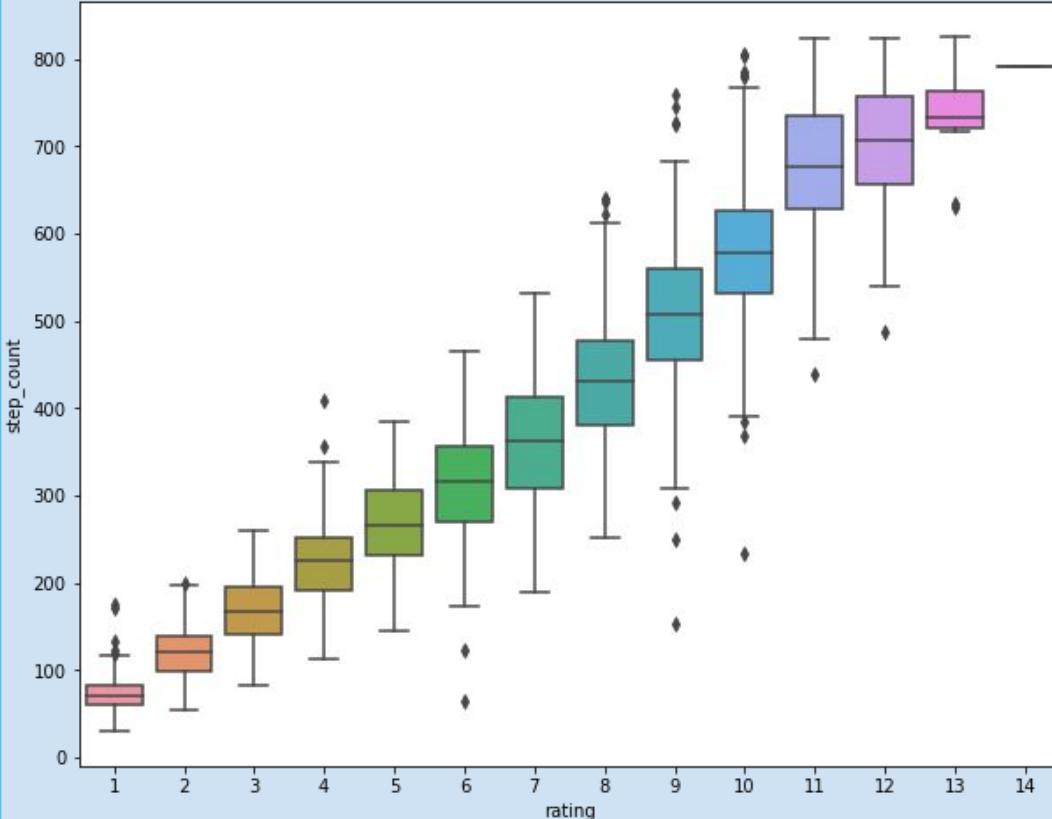
Key features: step count, song NPS, stream total

## Stamina

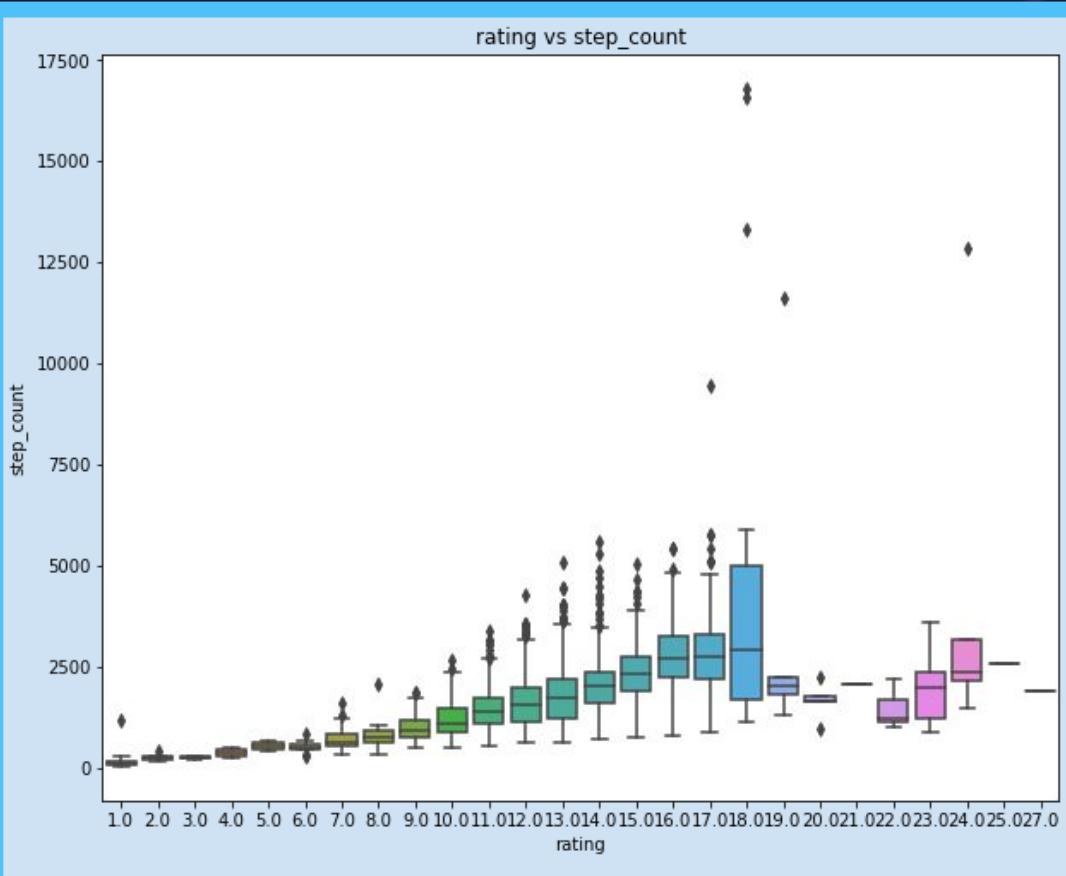
- Larger distribution - song lengths & difficulties
- Less pattern complexity

# Technical

rating vs step\_count



# Stamina



# First Simple Model

Tech  
Linear Regression

R-squared: .951

MSE: .642

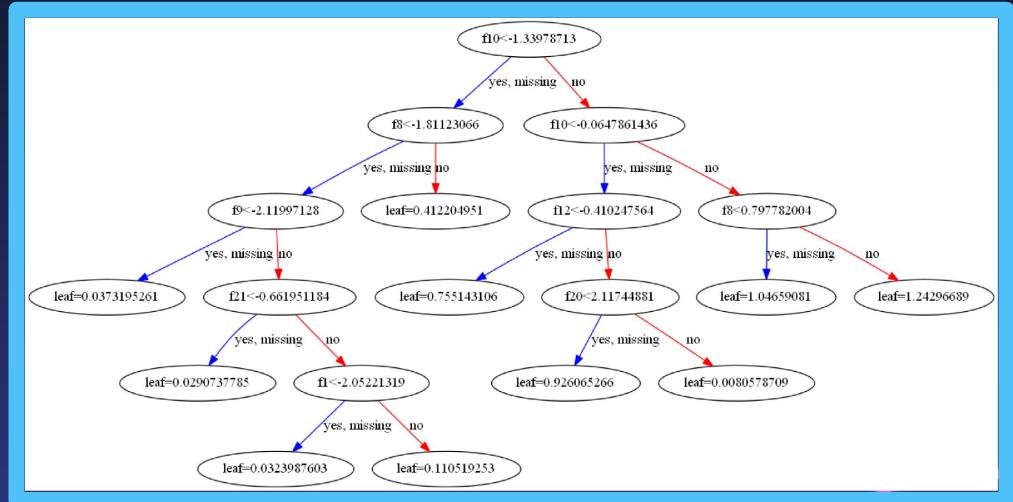
Stamina  
Linear Regression

R-squared: .958

MSE: .502

# XGBoost Regressor

- Tree-based Regression Model
- Boosts high-performing trees
- Decisions based upon residuals



# Final Model - Technical

## XGBoost Regressor

R-squared = .968

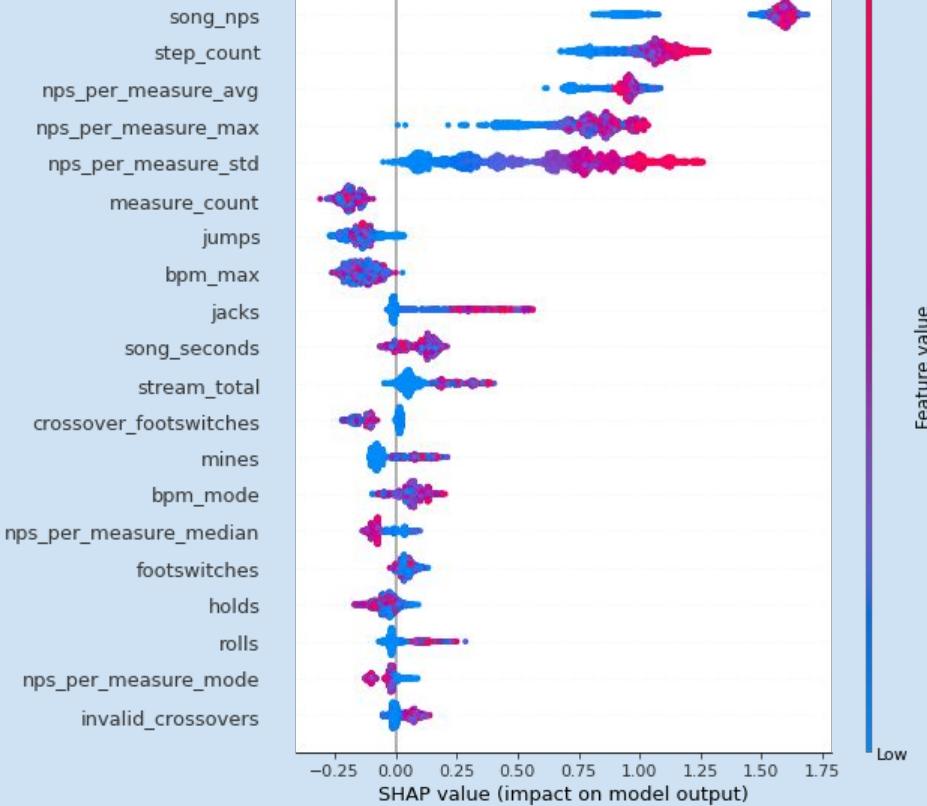
MSE = .404

learning rate = .1

depth = 5

gamma = 0

lambda = .1

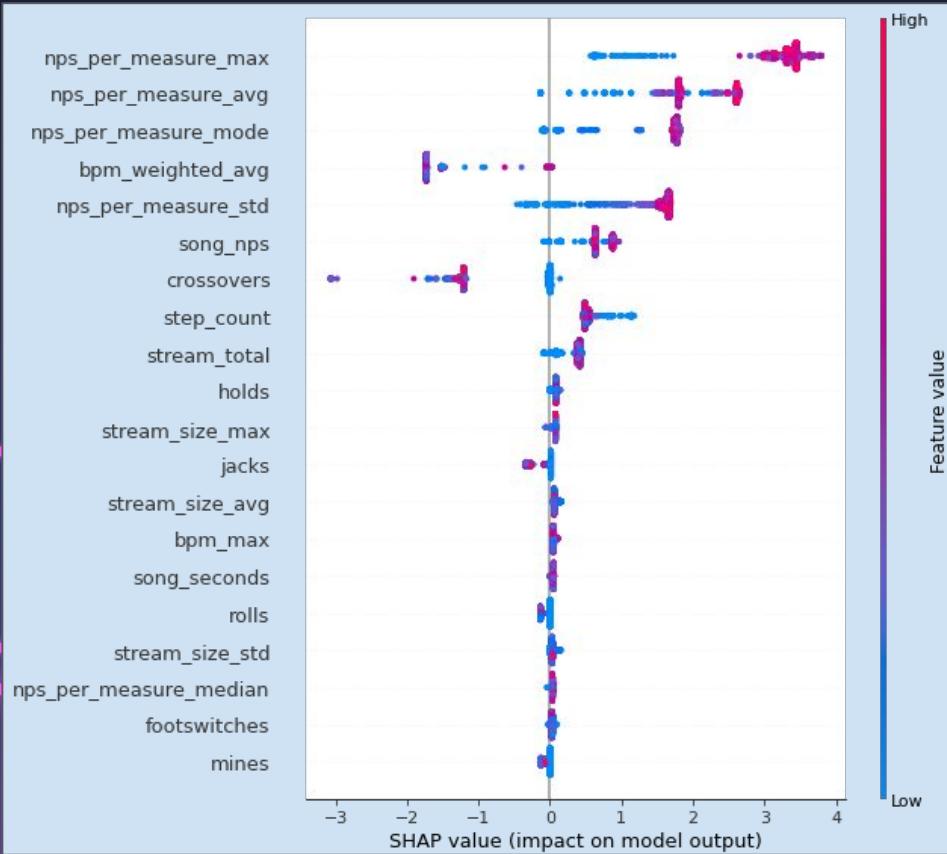


# Final Model - Stamina

## XGBoost Regressor

R-squared = .979  
MSE = .282

learning rate = .1  
depth = 5  
gamma = .5  
lambda = 5



# Next Steps



- Enable the user to upload pure song files to Flask app
- Tuning a classification model for Stamina vs. Tech
- Further data acquisition



# THANK YOU!

svitlana glibova

[s.glibova@gmail.com](mailto:s.glibova@gmail.com) |  | 

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.