

# Justin Chen

[chen.justi@northeastern.edu](mailto:chen.justi@northeastern.edu) • 650-388-0985 • <https://jyhchen.github.io/>

## EDUCATION

### Northeastern University, Khoury College of Computer Sciences

*Candidate for Bachelor of Science in Data Science and Business Administration*

**Awards and Honors:** Dean's Scholarship, Dean's List, University Honors Program

September 2019 - Present

Expected: Dec 2023

GPA: 3.51/4.0

**Coursework In:** Statistics; Object Oriented Design; SQL and no-SQL Information Storage; Supervised and Unsupervised Machine Learning; Data Visualization; Financial Data Analytics and Modeling; Sports Media, Marketing, and Communications  
**Activities:** Varsity Baseball Student Manager, WRBB Sports (student radio), Intramural Sports

## SKILLS AND INTERESTS

**Computer Languages:** Java, Python, R, Racket, HTML, CSS, Javascript

**Database:** SQL/mysql, Snowflake, MongoDB, Redis, Neo4j

**Frameworks and Packages:** Django, React.js, D3.js, AWS

**Software:** Git, IntelliJ, Jupyter, Adobe Analytics, Visual Studio Code, Tableau

**Interests:** Movie Soundtracks, Documentaries, Hats, Public Transportation, Biking

**Other:** Social Media, Sports Writing and Broadcasting, Chinese (native speaker)

## WORK EXPERIENCE

### Student Analyst and Manager

*Northeastern University Baseball*

Boston, MA

January 2020- Present

- Led team of managers, oversaw and planned analytics and data initiatives
- Created pitch and arsenal grading and comparisons with Rapsodo, YakkerTech, and Hawkeye metrics
- Reviewed offense strategy weekly with coaching staff to maximize value from lineups, bunts, and stolen bases
- Compiled scouting reports using statistics, play-by-play, and video from Synergy and TruMedia

### Web Analytics and A/B Testing Co-op

*Harvard Business Review*

Brighton, MA

July 2022- December 2022

- Created interactive reports with Adobe Analytics to summarize A/B test findings and analyze user engagement
- Designed experiments to analyze article recommendations, newsletter engagement, and SEO article performance
- Developed model to predict website visitors' conversion to paying subscriber with 80% accuracy
- Automated dashboard backup and creation with R and Python, utilizing Adobe APIs

### Baseball Systems Development Co-op

*Baltimore Orioles*

Baltimore, MD

July 2021- January 2022

- Rebuilt data pipelines for player stat views on the team's internal site with Django
- Developed amateur player evaluation forms and tables for scouts with React
- Designed quality checks and internal ID mapping for imported third-party data and created Slack notification bots
- Automated web scrapers with AWS lambda functions

## PROJECTS

### NCAA Basketball Prospect Ranker (Python/Jupyter Notebook)

- Collected dataset of 15,000 NCAA players with over 45 features and their NBA win shares with data scraping
- Used one-hot encoding and recursive feature elimination find the optimal feature subset
- Trained regression models with cross-validation, including Random Forest, AdaBoost, and Multilayer Perceptron
- Graded and ranked players for the 2022 Draft, with results similar to ESPN's projections

### Expected Run Value Model (xRV) and Pitch Grader (Python/Jupyter Notebook)

- Modeled expected run values to quantify pitch metrics, like movement, velocity, spin, in an aggregate
- Created pipelines for multi-parameter cross-validation and standardization to streamline preprocessing and training
- Analyzed feature importance and impact of different metrics to create the most comprehensive yet simplest models
- Graded MLB, WBC, and NCAA data to discuss the quality of pitches on Twitter and with the Northeastern coaching staff

### MLB Repository (mysql/mysql Workbench/Python)

- Created a relational database to store biographies and statistics for players and teams
- Implemented functions and triggers to update tables with additions or deletions of records
- Used Python to create a command-line application to access player and team information and statistics