Justin Chen

chen.justi@northeastern.edu • 650-388-0985 • https://jyhchen.github.io/

EDUCATION

Northeastern University, Khoury College of Computer Sciences September 2019 - Present
Candidate for Bachelor of Science in Data Science and Business Administration Expected: Dec 2023
Awards and Honors: Dean's Scholarship, Dean's List, University Honors Program GPA: 3.51/4.0

Coursework In: Statistics; Object Oriented Design; 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, Juptyer, 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 Boston, MA
Northeastern University Baseball 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 Brighton, MA
Harvard Business Review 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, MD
Baltimore Orioles 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

MLB Matchup Tool (Python/Jupyter Notebook)
- Cleaned pitch data and encoded categorical variables like balls, strikes, and pitch zones
- Grouped batters based on pitch and zone tendencies using K-Means to increase sample size for hitter-specific models
- Trained regression models with cross-validation, including Random Forest, XGBoost, and Ridge
- Created command-line application to give pitch recommendations taking for a pitcher against a specific hitter and count

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

AirBnb Recommendation Engine (Python/Neo4j/MongoDB)
- Used graph database in Neo4j to create recommendations based on quantitative features and node distances
- Implemented custom engine in Python to query Mongo for listings based on user specification and then an expand search
- Analyzed and discussed the advantages and shortcomings of each database at accomplishing similar tasks