

# Drew Szurko

[drewszurko.com](http://drewszurko.com) | [linkedin.com/in/drewszurko](https://www.linkedin.com/in/drewszurko) | [andrewszurko@gmail.com](mailto:andrewszurko@gmail.com) | (630) 346-7997

## EXPERIENCE

---

### Data/Machine Learning Engineer

*Self-Employed*

**February 2017 – Present**

*Chicago, IL*

- Utilized Spark and Kafka to develop a big data Python ETL pipeline for sentiment analysis.
- Designed SQL and NoSQL databases for mobile and desktop applications.
- Performed data pre-processing in Python to prepare datasets for statistical analysis.
- Built a XGBoost regression stack that finished in the top 4% of a Google Cloud machine learning competition.
- Leveraged Matplotlib to perform data analysis on large-scale datasets in Python.
- Developed Java and Kotlin Android applications that ranked top 20 and achieved over 40,200 downloads.

### Data Analyst

*Drop Tank, LLC.*

**May 2014 – February 2017**

*Burr Ridge, IL*

- Constructed an ETL pipeline to automate member activation using Java and MySQL.
- Saved 12 man-hours per week using JasperSoft and MySQL to automate daily reports.
- Provided conclusions to customers by analyzing, modeling and preparing reports from daily transaction data.
- Executed quality control checks during data migrations using custom SQL queries to ensure data integrity.
- Reduced SQL query execution times by up to 70% by consolidating redundancies and removing inefficiencies.

## EDUCATION

---

### Northern Illinois University

*B.S. Operations and Information Management*

**August 2009 – August 2013**

*DeKalb, IL*

## SKILLS

---

**Languages:** Python, SQL, NoSQL, Java, Kotlin, HTML, CSS, JavaScript

**Tools and Frameworks:** Git, Linux, Tableau, JasperSoft, Android, Django, Docker, Jira, Rasa, QGIS

**Technologies:** Spark, TensorFlow, Keras, Kafka, MySQL, PostgreSQL, HBase, LightGBM, XGBoost, OpenCV, scikit-learn, pandas, Matplotlib, NLTK, spaCy, NumPy, SciPy

## PROJECTS

---

### Twitter Sentiment Analysis

**June 2018 – August 2018**

- Developed a Python-Twitter ETL data pipeline to collect user tweets for sentiment analysis.
- Implemented Spark to calculate rolling sentiment averages on aggregated tweet data.
- Used Kafka to ensure high throughput and low latency when processing tweet topics.
- Extended pipeline functionality using HBase NoSQL database for storage of tweets.

### Daily Fantasy Sports NBA Machine Learning

**October 2017 – February 2019**

- Trained a Python machine learning regression model using historical NBA player stats.
- Leveraged LightGBM to predict player projections resulting in a 52.72% return on investment.
- Developed a custom Python OR-Tools algorithm to facilitate the building of bulk optimized lineups.
- Built Python-based ETL data collection web scraper for NBA games.

### Gemini AI

**December 2018 – Present**

- Created foundation for future Python natural language processing website using Django, Nginx, and Unicorn.
- Increased Keras sentiment analysis model accuracy by 3% by substituting a long short-term memory recurrent neural network with a n-gram multichannel convolutional neural network; reduced model train time by 60%.
- Implemented a Postgres database for storing of user objects and submissions.
- Improved accuracy of all Keras models by creating functions to remove stop words and perform lemmatization.

### Google Cloud & NCAA® ML 2018 (Top 4%)

**February 2018 – April 2018**

- Created a XGBoost gradient boosted machine learning ensemble of decision trees to predict win probabilities.
- Python model achieved a 0.574 log loss and placed in the top 4% of competing Kaggle teams.
- Reduced training log loss by over 25% using pandas for feature engineering.

### TensorFlow 2.0 WGAN-GP

**March 2019 – May 2019**

- Developed a Python TensorFlow 2.0 Wasserstein generative adversarial network.
- Integrated a gradient penalty to minimize vanishing/exploding gradients.
- Constructed a custom flexible big data pipeline to preprocess five different image datasets.