Thomas "Tommy" Zugibe Jr.

23 Fiske Avenue ✦ Somerville, Massachusetts 02145 ✦ 845-642-2423 ✦ Thomas.Zugibe@gmail.com

Summary

Tommy Zugibe is an accomplished and experienced data scientist with a strong background in developing machine learning models, troubleshooting and maintaining production-level pipelines, as well as utilizing statistical models to analyze and interpret data. He is a highly skilled problem solver, with a proven ability to develop end-to-end solutions for a wide range of data-driven problems, using a variety of scripting language software and tools, including Python, Django, PostgreSQL, Cassandra, and AWS. Tommy has worked in various industries, including health and wellness, network security, and transportation, developing and implementing solutions that have positively impacted business operations and user experience. He has experience as a Senior Data Scientist at Whoop, where he investigated the relationship between executive functioning test results and sleep recovery statistics, and as a Remote Principal Data Scientist at Plixer, where he developed project plans for network flow forecasting and malware-preventative security initiatives. In addition, he has acted as a Data Scientist at Agero, Inc., where he developed a handheld and hands-free phone activity detection and classification scheme. Tommy's career began as a Signal Processing Engineer at MIT Lincoln Laboratory, where he developed detection and classification algorithms for airborne radar systems. He is a talented collaborator and has acted as a team community, culture, and networking leader, planning team events, acting as the lead onboarding contact, and facilitating team morale.

Employment

- **Jellyfish,** Sr. Data Scientist   (Apr 2022 - Jan 2023)
  Jellyfish provides visibility into engineering organizations by analyzing engineering signals and contextual business data, helping engineering leaders align their work with strategic business objectives
    - Achieved an 18% business performance improvement in work hypothesis identification of our production core product, the Work Model, by extending the time window over which the Django query retrieves the computed work hypotheses in order to gain a better understanding of the actual probability distributions of work hypotheses over the processing period.
    - A new work version configuration was developed to gate the changes behind a feature flag for ease of rolling out the windowing buffer enhancements, as well as programming Django migration files to add new variables to our data tables
    - Perform multiple pipeline processing runs using Docker containers, EC2 instances, as well as staging and shell-plus environments to thoroughly test and validate the improvement of the time windowing buffers
    - Managed continuous improvement of the Work Model to score and display pull-request-related work events in real-time regardless of whether or not a PR has been successfully merged, as opposed to being surfaced posthumously only if it has been successfully merged. Ensured smooth rollout of new pipeline, associated migration files, and work version configurations through proper A/B testing experiments leveraging customer data.
    - Liaised with Sales Engineers and Customer Success Managers to advertise the new improvements and provided subject matter expertise to assuage any client fears during regularly scheduled customer-facing Gong calls.
    - Assumed the role of team culture and community leader by planning monthly team events, scheduling "water cooler" donut chats on Slack, onboarding new employees, and organizing team game nights to build morale.
    - Leverage customer data to perform a full data inventory of the company's available SQL-based ground truth data to facilitate future endeavors into machine learning model training improvements
    - Collaborate with a cross-functional team of six for the annual company hackathon and developed an award-winning Slack app to track company culture. Managers could use this tool to stay on top of employee morale, track sprint goals, balance meeting time with focus time, and assess general employee/company satisfaction NPS

- **Independent Author**   (Sep 2021 - Apr 2022)
    - Devoted full-time to writing a politically-based drama series in a fantasy setting, developing a unique world with distinct races, cultures, governments, and history
    - Maintained consistency throughout the series by paying meticulous attention to detail and research to create a believable and immersive world
    - Created multi-dimensional and relatable characters with unique personalities, motivations, and flaws, as well as established rules and limitations for the magic system through mastery of world-building, including its source, the types of magic, and the consequences of using it
    - Employed strong editing and revision skills to polish and refine the manuscript, remaining committed to completing the project

- **Whoop,** Senior Data Scientist   (Jan 2021 - Sep 2021)
  Whoop is a fitness and health monitoring company that tracks sleep, recovery, and daily effort to provide actionable insights on how to optimize performance, analyzing key metrics such as heart rate variability and resting heart rate to determine daily recovery score, and measuring training activities with a Strain score to help users understand when to rest or push themselves
    - Performed operations research to determine the relationship between users' N-Back and Stroop executive functioning (EF) test results to Whoop's daily "Recovery" statistic. Interpret results utilizing conventional covariance

metrics as well as a deeper unsupervised machine learning model to determine the efficacy of predicting a user's EF test scores from only our recovery metrics
- Developed an innovative and analytical Python-based library of troubleshooting tools to enable rapid testing of the various pipelines within the production code framework including data retrievals from Cassandra, PostgreSQL, and AWS
- Presented and documented the newly developed troubleshooting tools to the data science team to explain what is now available, how to install and use every script, as well as describe the architecture that enables the streamlining of new troubleshooting tools amidst an ever-changing codebase
- Represented the data science team for weekly on-call shifts aimed at triaging any bugs that arose and quickly acted to correct these issues to prevent any noticeable issues from reaching our ~100k active users
- Liaised with the Software Engineering, Data Engineering, and Coaching Infrastructure teams to utilize a newly developed heart rate confidence metric and achieved an accuracy improvement of 32% in our sleep-based recovery metrics including peak-to-peak pulse identification, heart rate variability, and respiration rate

- **Plixer,** Remote Principal Data Scientist   (Sep 2019 - Oct 2020)
Plixer provides deep network observability and intelligence platforms, including Plixer Security Intelligence and Plixer Network Intelligence, to detect and respond to cyber threats, predict and solve network interruptions, and gain visibility into network traffic and device behavior
    - Acted as leadership in the development of the complete project plan for machine learning integration for network flow forecasting as well as security initiatives to classify malware attacks
    - Collaboratively developed alongside business stakeholders, the entire network forecasting pipeline from ideation to Python productionalized real-time code modularized into the existing legacy Scala-written product
    - Single-handedly created the network classification framework utilizing a sophisticated Random Forest Classifier to provide both the first-ever automatic network device classification (WAN, router, personal computer, etc.) as well as adaptively detect any potential security threats to the system in real-time through Docker containers
    - Worked jointly in the maintenance and development of the team's datastream warehouse stored on our own Kubernetes server
    - Managed the work of a junior data scientist to facilitate optimal performance from the artificial intelligence team

- **Agero, Inc.**, Data Scientist   (Jun 2016 - May 2019)
Agero is a technology company that provides products and services to streamline and simplify roadside assistance, accident management, consumer affairs, and connected vehicle services
    - Development of an end-to-end handheld and handsfree phone activity detection and classification scheme through a Python-oriented adaptive noise floor cancellation process combined with a Dynamic Time Warping-based signal classifier
    - Demographic user behavior analysis as well as product analytics to categorize phone activity and driving trends of our users from the 1+ billion miles of data stored in our in-house and AWS Cloud-based data lakes (DynamoDB, Athena, and Redshift)
    - Integrated open-sourced municipal data directories to create a map-matching algorithm for transit mode detection in an effort to distinguish whether a user is a driver or a passenger
    - Employed numerous outlier and anomaly detection applications to identify fraudulent service providers
    - SQL database creation and management of Splunk stored survey logs to aid analysts' workflows
    - Led the service provider fraud endeavor by utilizing anomaly detection algorithms and saving ~$100k annually. Leverage customer data to create continuous improvement in the level of active fraud

- **MIT Lincoln Laboratory**, Signal Processing Engineer   (May 2010 - Jun 2016)
A federally funded research and development center (FFRDC) managed by MIT for the Department of Defense. Tasked with researching and developing advanced technologies for critical national security needs, including defending against missile threats, providing secure communications, monitoring space activity, and inventing biomedical devices, among others.
    - Algorithm development for airborne radar systems including image processing, classification, target tracking, and product experiments
    - Performance analysis and assessment of radar platform modes and applications
    - Regularized linear regression and logistical regression predictive analysis of ground/sea target classes
    - Machine learning applications to large data sets for event recognition and data tracking
    - Presented data findings and project concepts to the professional community

Education
- M.S. in Applied Mathematics, Rensselaer Polytechnic Institute, Troy, New York, 2010, GPA: 4.00
- B.S. in Applied Mathematics, Marist College, Poughkeepsie, New York, 2008, 3.92

Skills
- Flexible, positive, and productive when it comes to collaboration
- Assertive without being competitive in the decision making process through the use of problem-solving skills alongside a measure of accountability for said actions
- Experienced Business Intelligence (BI) analysis with a focus on data visualization using software such as Tableau

- In-depth background in math, statistics, economics, and financial modeling
- Dedicated to continuous improvement of myself with a willingness to learn
- Comfortable working remotely, in a satellite office, but enjoy networking with colleagues in a more hybrid environment

Awards, Fellowships, Grants
- Awarded Most Marketable Hackathon Project, Jellyfish (2022)
- Awarded MV^2 (Most Valuable Vertical), Agero (2017 & 2018)
- Awarded Research Assistantship, Rensselaer Polytechnic Institute (2009)
- Awarded Teaching Assistantship, Rensselaer Polytechnic Institute (2008-2009)
- Single Baccalaureate recipient in Mathematics, Marist College (2008)
- Summa Cum Laude, Rensselaer Polytechnic Institute (2010)
- Magna Cum Laude, Marist College (2008)
- Presidential scholarship for advanced academic achievement, Marist College (2004-2008)
- Graduate of the prestigious Honors Program, Marist College (2008)
- Graduate with honors in Mathematics, School of Computer Science and Mathematics, Marist College (2008)
- Grant to conduct independent research on seismological systems, School of Computer Science and Mathematics, Marist College (2007)