

# ROHAN MANOJ THAKKAR

[rohan.m.thakkar@gmail.com](mailto:rohan.m.thakkar@gmail.com)  
[linkedin.com/in/rohanmthakkar](https://www.linkedin.com/in/rohanmthakkar)

Seattle, WA  
[27rohan.github.io/Portfolio](https://27rohan.github.io/Portfolio)

(206)739-4797  
[github.com/27rohan](https://github.com/27rohan)

## EXPERIENCE

**TWITTER**, Seattle, WA

*Senior Machine Learning Engineer (Feb '22 - Present)*

- Launched cloud-based **continuous training & evaluation** pipelines for **Recommended Notifications**
  - Time to **train** neural network-based prediction models reduced by 95% (24+ hours to 1 hour)
  - Time to **evaluate** neural network-based prediction models reduced by 80% (4+ hours to 45m)
  - Online performance improved by 25% within 1-month of launch (PR-AUC 0.35 to 0.45)
  - Online traffic **serving** latency reduced by 90% (100-200 ms to 5-10 ms)
  - Number of oncall incidents reduced by 66% (3x/week to 1x/week)
- **ML Technologies:** Python, Kubeflow, Tensorflow (w/ TFX & Tensorboard), Airflow, GCP, OnPrem
- Tweaked **30+** data pipelines (100GB - 5PB/hour range) for performance, costs & external dependencies
  - 3 feature aggregation jobs for **Push Notifications**: Costs down 60% & Speed up by 50%
  - 2 feature generation pipelines for **Recommended Notifications**: Transitioned from Dataflow to Flink
  - 1 **Trends** data pipeline transitioned from GCP (Bigquery) to OnPrem (SparkSQL): Costs down 45%
  - 5+ **Client Log** databases transitioned from LZO to parquet format, with SparkSQL support
  - 20+ **Ads** data pipelines transitioned from GCP (Bigquery/Scalding) to OnPrem (SparkSQL/Scalding)
- **Data Eng Technologies:** Scala, Spark, Hive, Scalding, Beam, Dataflow, Flink, Airflow, GCP, OnPrem
- Mentored 3 FT engineers (SWE 1, SWE 2, Sr. MLE) & managed 1 SWE intern

**ZILLOW GROUP**, Seattle, WA

*Machine Learning Engineer (Jan '20 - Feb '22)*

- Created a **marketing engine** to automate multivariate experimentation & optimization across ZG
  - Onboarded two marketing campaigns for a collective annual revenue of \$15M+
- Collaborated with scientists, engineers & PM to launch ML models at 100M+ user scale
  - Time to launch new ML models reduced by 85% (2 weeks to 2 days)
  - Time to retrain new ML models reduced by 85% (1 weeks to 1 day)
  - Time to launch offline inference & validation pipelines reduced by 99% (1 week to 1 hour)
- Created an ingestion service to export all user scores & experimentation decisions to 4 ZG platforms daily
- Launched & maintained AWS **infra** for a team of 15+ to prototype & productionize big data workloads
  - Sustainable process for creating & using custom AMIs to save ~12% in AWS costs (~50k/year)
  - Automated backup & restore to prevent deletion of work saved on EMR during restarts
  - Jupyter notebook launching script with preinstalled packages & users tagged on resource manager
  - AWS roles with necessary datalake permissions, cluster monitor, Git support & more
- Added features for **Facebook Ads** Insights ETL & cut down FB dynamic costs by 30% via dynamic dedup
- Kickstarted & led team's **oncall** processes, including weekly handoff meetings & grooming OPS tickets
- Led ~50% of team's **Sprint** planning & **Retro** w/ stakeholder communication for key timelines & risks
- Mentored 2 FT peers (MLE & Applied Scientist) & 1 summer intern (eventually FT Applied Scientist)
- Authored [Zillow 2020 AI Forum recap](#)

**Technologies:** Airflow, Python, Spark, AWS, Hive, Terraform, Jupyter, Facebook Business

**VMWARE**, Palo Alto, CA

*Member of Technical Staff (Jun '17 - Jan '20)*

- Launched data pipelines, reporting, analytics & predictions for insights into engineering productivity metrics
- Mentored 2 FT engineers & 1 summer intern (eventually FT engineer)

**Technologies:** Python, Bash, SQL, NoSQL, YAML, Tableau

**CYANOGEN**, Seattle, WA

*Software Engineering Intern (Jun - Dec '16)*

- Co-developed 2 cloud-based ML services (App Recommender & NewsCard ranking)
- Created an e2e log analytics engine w/ online dashboarding support for engineering visibility

**Technologies:** AWS (Kafka, S3, TitanDB), Python, Bash, D3.js, HTML, CSS

# ROHAN MANOJ THAKKAR

[rohan.m.thakkar@gmail.com](mailto:rohan.m.thakkar@gmail.com)  
[linkedin.com/in/rohanmthakkar](https://www.linkedin.com/in/rohanmthakkar)

Seattle, WA  
[27rohan.github.io/Portfolio](https://27rohan.github.io/Portfolio)

(206)739-4797  
[github.com/27rohan](https://github.com/27rohan)

## EDUCATION

---

**M.S.,** University of Washington, Seattle, WA

*Information Management (Sep '15 - Jun '17)*

- Specializations in Data Science, Business Intelligence & Mobile App development

**B. Eng.,** University of Mumbai, India

*Computer Engineering (Aug '11 - May '15)*

- Capstone project in analyzing user behavior in e-commerce

## CERTIFICATIONS

---

**Nanodegree,** Udacity, Palo Alto, CA

*Intro to Self-Driving Cars (Mar - Jun '19)*

## PUBLICATIONS

---

**Business Analytics in Traffic Management System,** International Journal of Computer Applications, Volume 126 - No.12, September 2015

**Revamping Quality of Service of Video Streaming over Wireless LAN,** International Journal of Research in Engineering and Technology, Volume: 03 Issue: 07 | Jul-2014

## COMMUNITY SERVICE

---

**Computer Literacy Project,** Citizens Association for Child Rights

**HIV/AIDS Awareness in Migrant Workers,** Yuvak Pratishthan