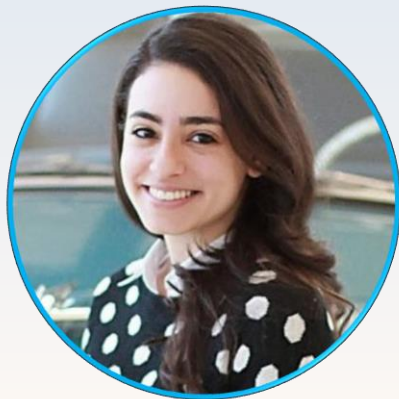


WELCOME



UNIVERSITY PREDICTION HACKATHON

Meet Allstate's Hackathon Team



LAURA FORMAN

Data Science
Recruiter



SARA RING

Data Engineer &
Hackathon Coordinator



SEAN WALSH

Hackathon
Communications



JEREMY WERNER

Data Scientist &
Hackathon Lead



... About Us ...

About ALLSTATE

FACTS & FIGURES

- Largest public personal lines insurer in the US
- We serve 16 million households
- We have 11,000 agencies and 40,000 employees supporting our customers

CORE VALUES

Diversity, balance, innovation and providing an environment where employees can thrive.
These are all core values at Allstate.

AWARDS

- The World's Leading Top 100 Companies – Forbes
- America's Best 50 Companies for Minorities – Fortune
- 50 Happiest Companies in America – CareerBliss
- Best Places to Work for Recent Grads – ConnectEDU
- ...just to name a few!

About DATA SCIENCE AT ALLSTATE

WHAT WE DO:

AGENCY ANALYTICS
TELEMATICS FOR INSURANCE
FRAUD DETECTION
CUSTOMER RETENTION
PRICING STRATEGIES

OUR TEAM:

50+
PhDs

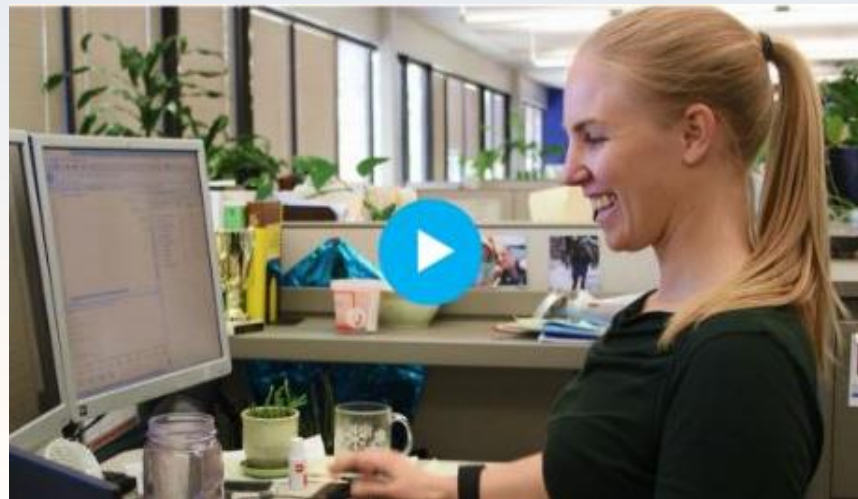
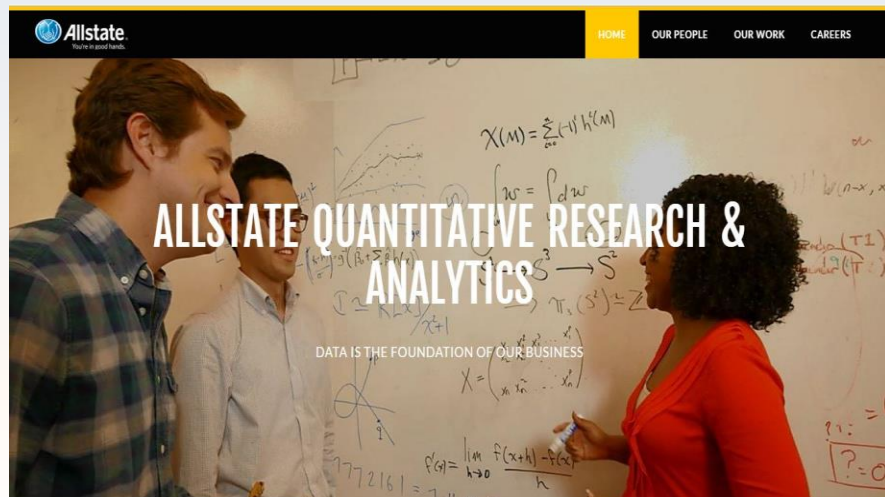
90+
master's
degrees



We have data scientists in five locations
Northbrook, IL • Chicago, IL • Seattle, WA
Silicon Valley • Belfast, Northern Ireland



Learn more about us at AllstateDataScience.com



Insurance Customer Event Prediction

Insurance carriers gain a competitive advantage from better understanding their customers & predicting their next likely event.

Example 1:

Open Policy -> Add vehicle -> Receive Bill -> Pay bill -> Renew Policy -> Add second vehicle -> ??

Example 2:

Suppose an insured customer changes their marital status from single to married.

- The insured could benefit from adding their new spouse to the account.
 - Does their new spouse also have a vehicle or home to be insured?

Understanding these patterns and leveraging them is mutually beneficial for Allstate and our customers.



PROBLEM STATEMENT:

“Life events” happen often.
Given an ordinal sequence of
customers’ life events, can we predict
what event will come next?



CHALLENGE:



Predict the next event in the
sequence



The Data

- train.csv

id	timestamp	event
00005408fb82819bf6eef036180ce1d28ca64ac455378824a882a80ed8f99bb6	1	36003
00005408fb82819bf6eef036180ce1d28ca64ac455378824a882a80ed8f99bb6	2	36003
00005408fb82819bf6eef036180ce1d28ca64ac455378824a882a80ed8f99bb6	3	30018
00005408fb82819bf6eef036180ce1d28ca64ac455378824a882a80ed8f99bb6	4	36003
000061e45fb216f4ad7fbc0cd86f620441a3091005eeb6fc76e1f71fdde1ecbc	5	30021
000061e45fb216f4ad7fbc0cd86f620441a3091005eeb6fc76e1f71fdde1ecbc	6	30027
000061e45fb216f4ad7fbc0cd86f620441a3091005eeb6fc76e1f71fdde1ecbc	7	30042
00008faca7acd5b2edf91b274eedc88e90b1de3b4003f93b67c20d425a29ee8d	8	30024
00008faca7acd5b2edf91b274eedc88e90b1de3b4003f93b67c20d425a29ee8d	9	30039
00008faca7acd5b2edf91b274eedc88e90b1de3b4003f93b67c20d425a29ee8d	10	30018

- test.csv

id
0001da55d168196bf25f06a497b5cf414126542f4d357d61b8203d3b2c9eb4e5
00024eca1053d4268df5c6d3308f0d008cf5e2678c443f46f83dd8a7714462c6
000273e55809afd4a9ac4fb9175effe5d0ea449ed37e7c924af3f993b2d3e14b
0002bd1d73c326ad6e337a5687f6787b055f13079d5c52ef5d21321ea233fa34
00030b6d5b8013bcb9bb23a9bccf394d7b361a01d0d6b70de5ffa99786d49b06



The Data

- **train.csv**
 - **Columns**
 - **id**: Unique identifier (string)
 - **timestamp**: Global timestamp (int)
 - **event**: Event code, 10 total (int)
 - **84 MB**
 - **1,122,007 rows**
- **test.csv**
 - **Column**
 - **id**: Unique identifier (string)
 - **6.3 MB**
 - **100,870 rows**



Submission Format

Submission needs to be split into columns by event probability

id	event_30018	event_30021	event_30024	event_30027	event_30039	event_30042	event_30045	event_30048	event_36003	event_45003
0001da55c	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
00024eca1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
000273e55	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0002bd1d	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1



THE VALIDATION METRIC

Multiclass Log Loss =

$$-\frac{1}{N} \sum_i \sum_j y_{ij} \log_e \hat{p}_{ij}$$

For all observations $i = 1, \dots, N$

For all classes $j = 1, \dots, k$

Lower is better.



A FEW NOTES ABOUT SUBMISSIONS

- ✓ Comma separated
- ✓ Eleven columns: "id", "event_30018", ..., "event_45003"
- ✓ All column values (other than id) are numeric between 0 and 1
- ✓ File should have 100870 predictions + 1 header row = 100871 rows

<https://inclass.kaggle.com/c/allstate-event-prediction-challenge>



Team Names

The team name must designate the university's initials followed by a single dash and a unique name

Examples:

- UD – Modeling Dynasty
- NU – 1NonLinearGal

University	Initials for Competition
Miami University of Ohio	MU
Northwestern University	NU
Stanford University	SU
University of California Davis	UD
University of Chicago	UC
University of Michigan	UM
University of Washington	UW



Leaderboard

- Multiclass Log Loss on the public leaderboard data for your best submission thus far is shown to all competitors throughout the competition.
- <https://inclass.kaggle.com/c/allstate-event-prediction-challenge/leaderboard>
- A benchmark submission is shown on the leaderboard for you to beat.



A FEW NOTES ABOUT RULES

PARTICIPANTS



- Contestants are not permitted to collaborate as a team. Each Contestant must register individually for the Contest using their university email address.
- Entrance into the competition is not limited to any time frame, but submissions must happen before the deadline.

RESOURCES



- Competitors use computing hardware that they provide
- Recommended (free) modeling software:
 - R
 - Python
- Use of proprietary software could lead to disqualification.

WINNER REQUIREMENTS



- Top scoring teams on the private leaderboard will be emailed at the email addresses corresponding to their Kaggle in Class accounts.
 - Timely response necessary
- Code will be reviewed by Allstate data scientists to ensure it reasonably created the supplied predictions and is in compliance with rules (e.g. no external data)
 - Keep code clean, cohesive, coherent, and saved!!



Prizes



\$1,000 will be awarded to the first place student.



\$500 will be awarded to the second place student.



\$100 will be awarded to top student at each school
(not already in first or second place overall)



Competition Resources

Forum:

Allstate data scientists will be policing the forums. Ask your questions there!

<https://inclass.kaggle.com/c/allstate-event-prediction-challenge/forums>

Tutorials:

Two tutorials have been prepared to help you get started and reproduce the baseline submission.

- R
- Python



ALLSTATE'S UNIVERSITY HACKATHON COMPETITION

1

Contest Information Website

- [AllstateDataScience.com/hack](https://allstatedata.com/hack)
- Tutorials for getting started on kaggle

2

Registration, Data, Submission, Leaderboard, and Forum site:

- <https://inclass.kaggle.com/c/allstate-event-prediction-challenge>

3

Other Items

- Contest closes Sunday, February 26th, **11:59:59 a.m. CT**
- Awards mailed to students
- Official contest rules on Kaggle
- hack@allstate.com

