

Intuit Data Classification Challenge



Intuit for Startups
July, 2024

INTUIT



turbotax



credit karma



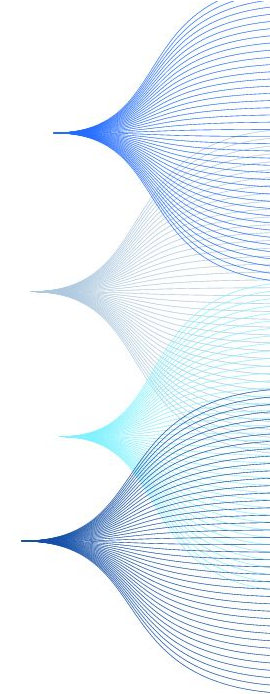
quickbooks



mailchimp

Webinar agenda

1. Introducing Intuit and Intuit for Startups
2. Why is Intuit investing in the challenge?
3. Challenge eligibility and fit criteria
4. Challenge timeline
5. Classification example
6. Scoring & guidelines
7. Live challenge technical overview
8. Reward
9. Q&A



Introducing Intuit

Intuit is the global financial technology platform that powers prosperity for the people and communities we serve



Smartest business tools for the world's hardest workers



Everything you need to get your taxes done right



Free Credit score and Credit reports



All in one marketing platform

Introducing Intuit

Intuit is the global financial technology platform that powers prosperity for the people and communities we serve

1983

Founded

1993

IPO

18,000+

Employees

19

Locations

\$14.4B

FY '23 Revenue

100M+

Customers

Intuit for Startups

Where Intuit and startups meet to power prosperity for customers and small businesses around the world.

We accelerate innovation at Intuit through:

Partnerships



Investments



Challenges



Acquisitions



Why is Intuit investing in the challenge?

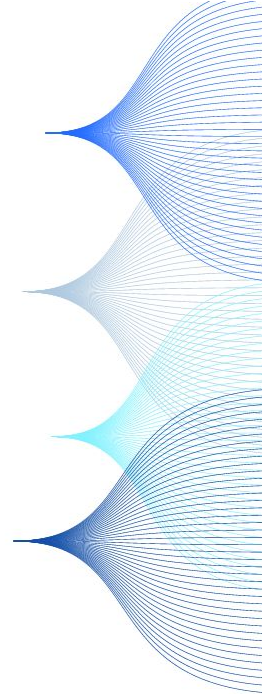
Our users entrust us with their most sensitive data

If every Intuit user only has 5 personal records

Then we would need to classify and protect **500,000,000** records

Our users have ***a lot more than just 5 records***

So we protect a lot more than 500,000,000+ records



Why is Intuit investing in the challenge?

When our users think of Intuit, they think of trust

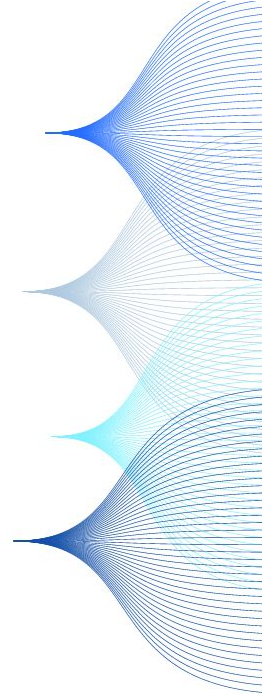
But *bad actors think of endless opportunities*

They're are using GenAI for bad

We built Intuit Assist for good and for the good of our users

Protecting data is the foundation for the Intuit of today and tomorrow

That's why investing in earning our users' trust is a strategic priority



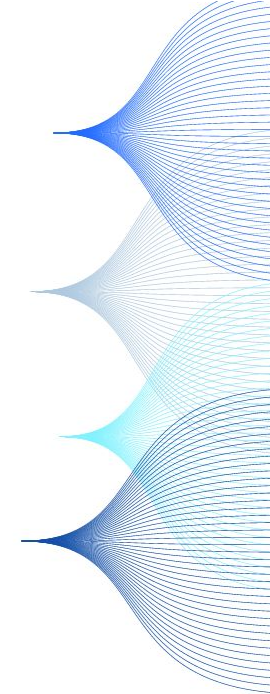
Challenge eligibility and fit criteria

Eligibility:

- Incorporated in the US or Israel
- Main line of business is a tech based data classification solution that has customers
- Completed application form by August 15th

Fit:

- Capable of supporting the scale of Intuit
- Cutting-edge data classification capabilities



Challenge timeline



Jul 24-25th

Challenge
Webinars



Aug 15th

Application
Deadline



Aug 26th

Application
Decision Sent



Sept 4th

Participants
Receive
Training Data
Set



LIVE

Sept 18th

Live Challenge



October

Winner
Announced

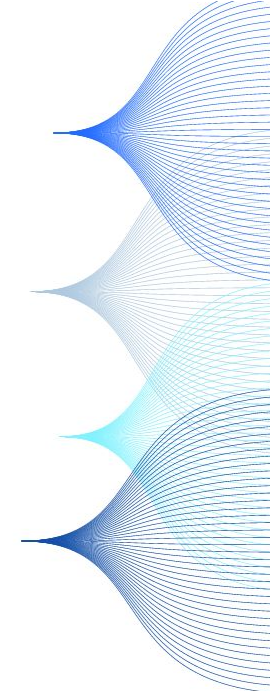
Classification Example

“SSN# Julie 309 12 3456 Ben
309-54-3216 4/10/23 D.O.B. Julie
4APR99 Ben birthday 3/14/95
2/10/23 I have Type-1 Diabetes”

Classification Example

“SSN# Julie 309 12 3456 Ben
309-54-3216 4/10/23 D.O.B. Julie
4APR99 Ben birthday 3/14/95
2/10/23 I have Type-1 Diabetes”

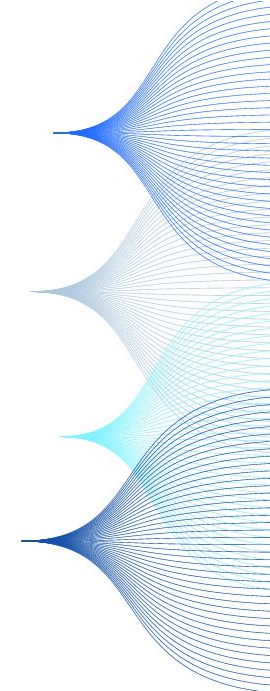
- Person Name



Classification Example

“SSN# Julie 309 12 3456 Ben
309-54-3216 4/10/23 D.O.B. Julie
4APR99 Ben birthday 3/14/95
2/10/23 I have Type-1 Diabetes”

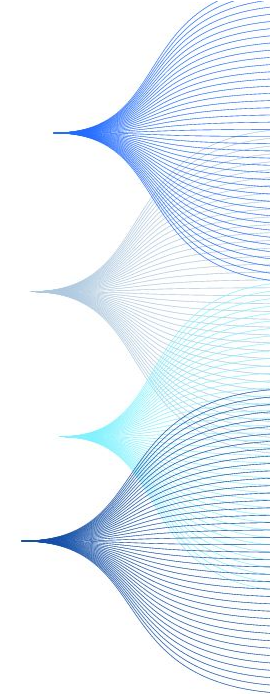
- Person Name
- SSN



Classification Example

“SSN# Julie 309 12 3456 Ben
309-54-3216 4/10/23 D.O.B. Julie
4APR99 Ben birthday 3/14/95
2/10/23 I have Type-1 Diabetes”

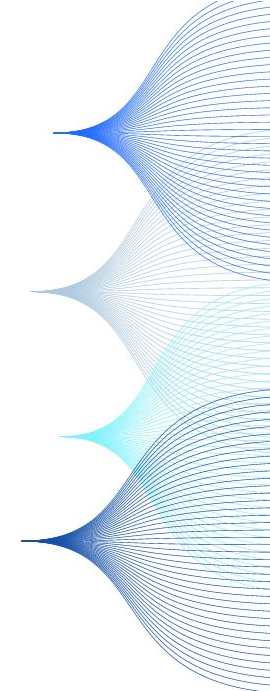
- Person Name
- SSN
- Date of birth



Classification Example

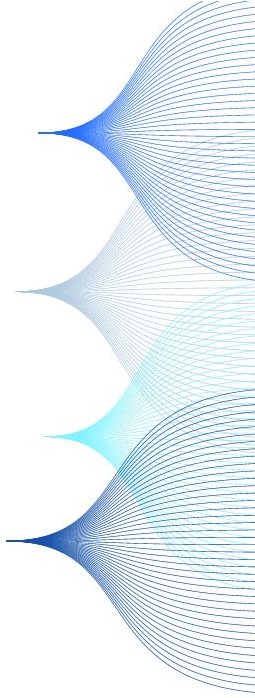
“SSN# Julie 309 12 3456 Ben
309-54-3216 4/10/23 D.O.B. Julie
4APR99 Ben birthday 3/14/95
2/10/23 I have Type-1 Diabetes”

- Person Name
- SSN
- Date of birth
- Medical Condition



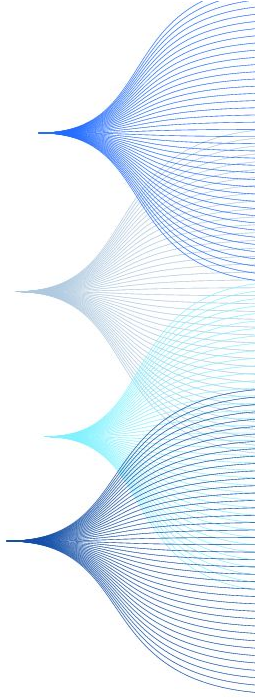
Classification Answer Structure

```
{'text': "SSN# Julie 309 12 3456 Ben  
309-54-3216 4/10/23 D.O.B. Julie 4APR99 Ben  
birthday 3/14/95 2/10/23 I have Type-1  
Diabetes",  
'detections': [  
  {'value': "Julie", 'start_index': 5,  
    'end_index': 10, 'label': 'Person Name'},  
  {'value': "309 12 3456", 'start_index': 11,  
    'end_index': 22, 'label': 'Social Security  
Number (SSN)'}, ... ]}
```



Classification Answer Structure

```
{'text': "SSN# Julie 309 12 3456 Ben  
309-54-3216 4/10/23 D.O.B. Julie 4APR99 Ben  
birthday 3/14/95 2/10/23 I have Type-1  
Diabetes",  
'detections':[  
{'value': "Julie", 'start_index': 5,  
'end_index': 10, 'label': 'Person Name'},  
{'value': "309 12 3456", 'start_index': 11,  
'end_index': 22, 'label': 'Social Security  
Number (SSN)'}, ... ]}
```



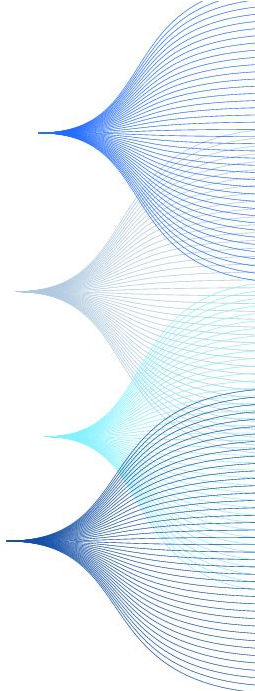
Classification Full Answer

```
{'text': "SSN# Julie 309 12 3456 Ben 309-54-3216 4/10/23 D.O.B. Julie 4APR99 Ben  
birthday 3/14/95 2/10/23 I have Type-1 Diabetes",  
'detections':  
[{'value': "Julie", 'start_index': 5, 'end_index': 10, 'label': 'Person Name'},  
{'value': "309 12 3456", 'start_index': 11, 'end_index': 22, 'label': 'Social  
Security Number (SSN)'},  
{'value': "Ben", 'start_index': 23, 'end_index': 26, 'label': 'Person Name'},  
{'value': "309-54-3216", 'start_index': 27, 'end_index': 38, 'label': 'Social  
Security Number (SSN)'},  
{'value': "Julie", 'start_index': 54, 'end_index': 59, 'label': 'Person Name'},  
{'value': "4APR99", 'start_index': 60, 'end_index': 66, 'label': 'Date of Birth  
(DOB)'},  
{'value': "Ben", 'start_index': 67, 'end_index': 70, 'label': 'Person Name'},  
{'value': "3/14/95", 'start_index': 80, 'end_index': 87, 'label': 'Date of Birth  
(DOB)'},  
{'value': "Type-1 Diabetes", 'start_index': 103, 'end_index': 118, 'label':  
'Medical Condition'}  
]}
```

Scoring

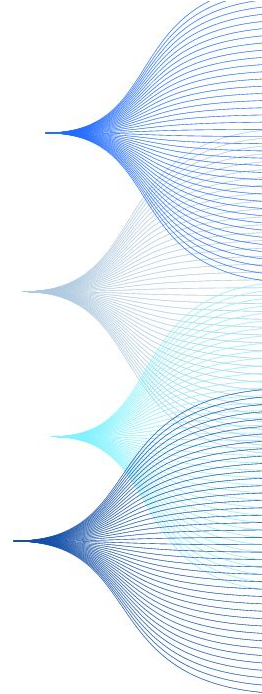
- Each entity is evaluated separately (Recall and Precision)
- We take into account partial correct matches, although it will lower the score.
- Overall performance is a weighted average of all the entities

Weight 40%	Weight 30%	Weight 20% Default	Weight 10%
Mexican Population	Card Verification	SSN	Mexican Population
The complete list will be available on the microsite			
Username			Number
Driver License	Bank Account	Credit Card Expiration Date	Email Address



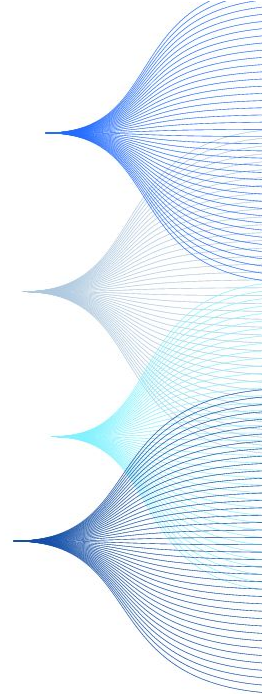
Guidelines

- Unstructured and structured data
 - Free text
 - JSON
- Validating the values (examples)
 - Credit Card Number is valid only if it passes Luhn check
 - SSN can not begin with '9'
- Multiple possibilities for one value
 - Choose the one with highest confidence
- Multiple languages (English, Spanish, Portuguese).
- Have a question?
 - Please send the question to Intuit4Startups@intuit.com



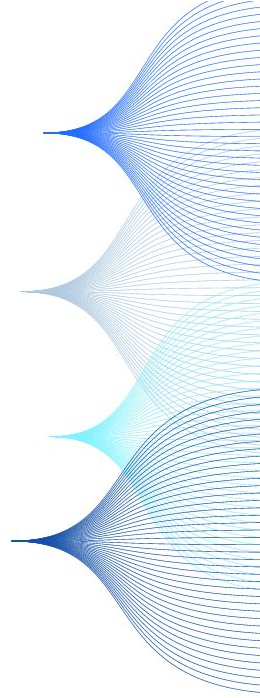
Live challenge technical overview

- Training and Challenge data will be shared via Google Drive.
 - Each participant will share email to get access.
- The training data ~20k
 - Text with labeled entities
 - JSONL (list of json objects)
- The challenge data ~20k
 - Text without labels (same distribution as training data)
 - JSONL (list of json objects)
 - Solutions will be uploaded to Google Drive location per participant.
- The challenge
 - 20 min to classify and upload your solution.
- We will be available on zoom call to assist.



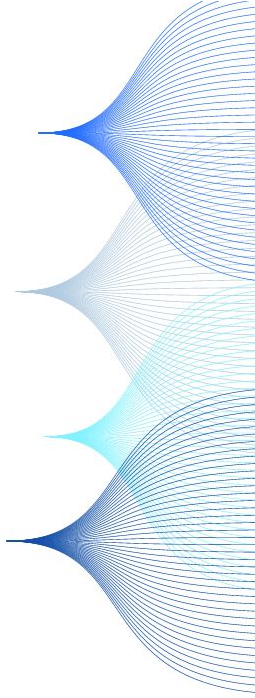
Reward

- The winning team will present their technology and solution to senior Intuit executives
- A prize of \$10,000
- *This is not an RFI or RFP*
- **It is a unique opportunity to:**
 1. Build a relationship directly with Intuit executives
 2. Receive objective public recognition as market leaders by Intuit

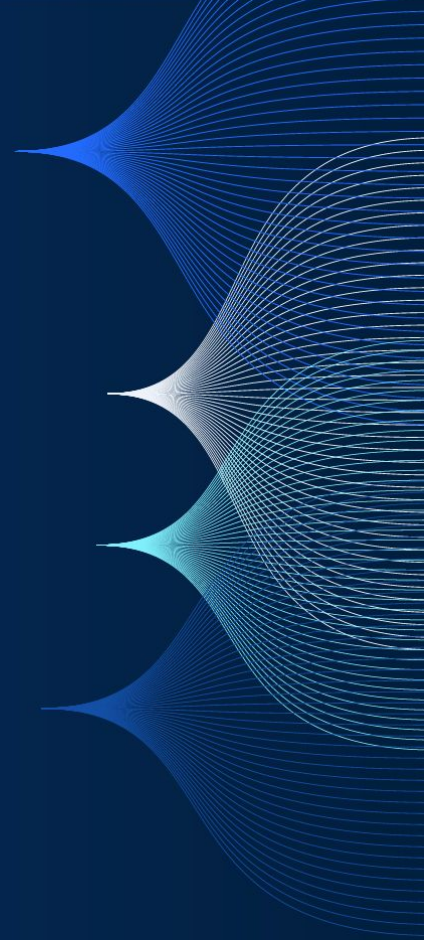


Q&A

- We'll start with the questions you submitted
- Raise your hands and we will call on you



Thank you and
good luck!



INTUIT



turbotax



credit karma



quickbooks



mailchimp