dotscience

Chaos and Pain in Machine Learning, and the ‘DevOps for ML Manifesto’

Mark Coleman, VP Product & Marketing at Dotscience

AI has the potential to reinvent the global economy

But as a discipline it's immature — **it's the Wild West out there**

We've seen damaging levels of chaos and pain operationalizing AI

2019 State of Development and Operations in AI report

Surveyed 500 companies which confirmed that it's **definitely** the wild west out there

Download the survey report at:

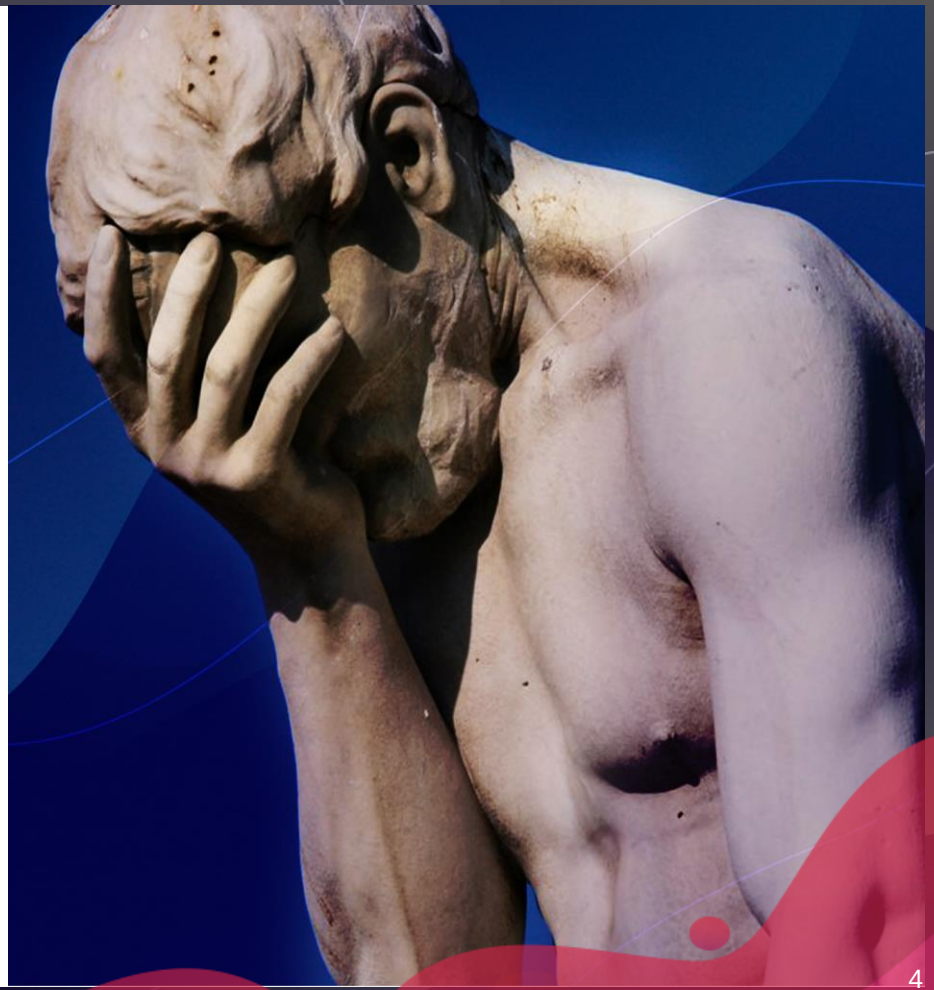
dotscience.com/2019-survey/

Key problems affecting AI efforts today

Many teams have problems they don't realize exist, because it's just "how it's done"

1. Wasting time
2. Inefficient collaboration
3. Manual tracking
4. No reproducibility or provenance
5. Unmonitored models

and more...



dotscience

Key survey findings

While 62.3% of businesses reported they are spending between $500,000 and $10,000,000 on their AI efforts, 60.6% of respondents are continuing to experience a variety of operational challenges

The top three challenges respondents experienced with AI workloads are duplicating work (33%), rewriting models after team members leave (27.6%) and difficulty justifying value (27%)

64.4% of respondents take between 7 and 18 months to move ML models from idea to production

44.5% of ML engineers and data scientists collaborate with each other using a shared spreadsheet for metrics which they update manually

dotscience

Quotes from the frontiers...

In retrospect if we had been able to save the versions or have gone back in time to see how he got his learning rates it would have avoided a lot of questions from the auditors.

Two of the data scientists who worked on that particular model have left and gone to other companies. You want to be able to see what they did and how they did it and not that it's gone when they're gone.

“

One model failed for 3 months and we lost an immeasurable amount of money!

After the last audit I was surprised by how many problems in the audit we could have solved by keeping PAPER LOGS. But if we ask our data scientists to do this they will leave!

We keep our data science teams small and in the same room so they can track their summary statistics by talking to each other and remembering

We've been here before!

In the 90s, software engineering was siloed.

Not everything was version controlled, there was no continuous delivery.

Software took months to ship; now it ships in minutes.

The same is possible for AI...

Requirements to achieve DevOps for ML

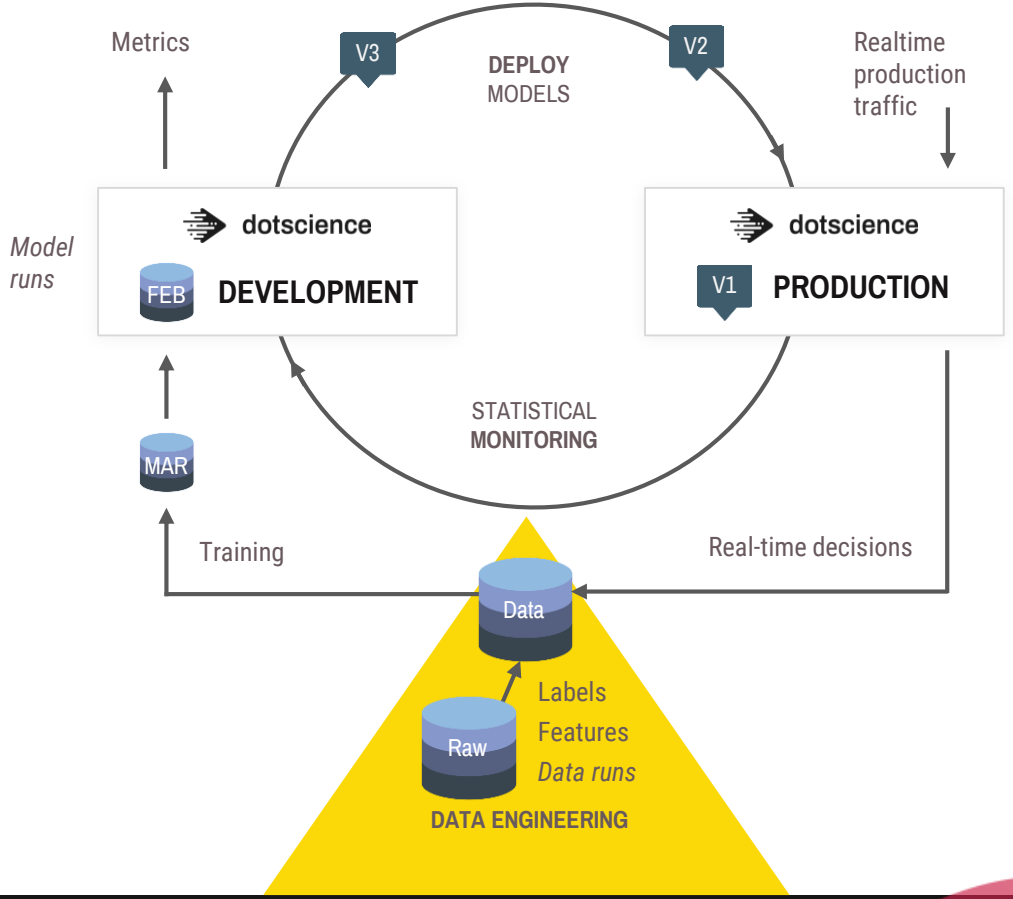
1. Reproducible
Must be able to retrain a 9 month old model to within few %

2. Accountable
Must be able to trace back from model in production to its provenance

3. Collaborative
Must be able to do asynchronous collaboration

4. Continuous
Must be able to deploy automatically & monitor statistically

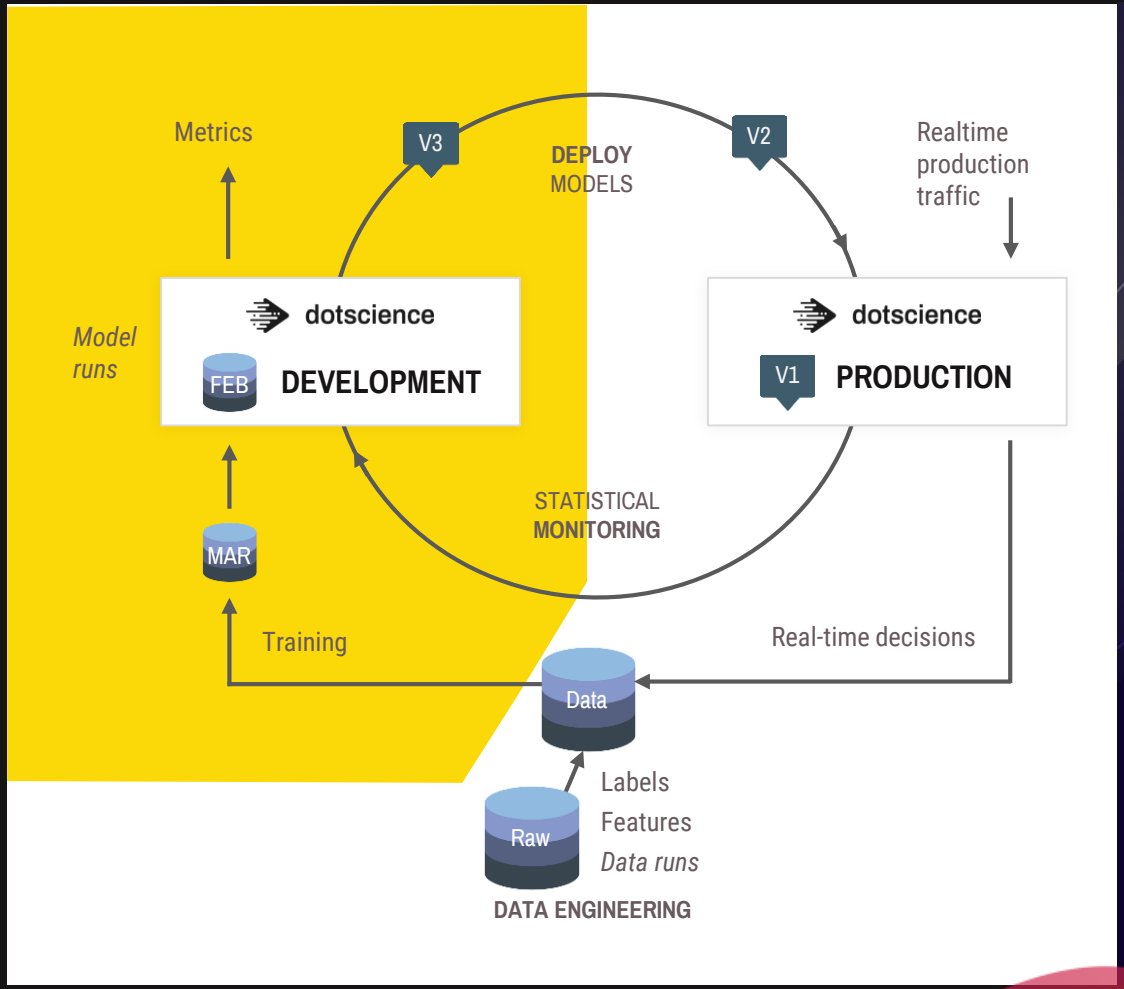
The model lifecycle.



Data Engineering

In data engineering pipelines, you need to track *data runs*.

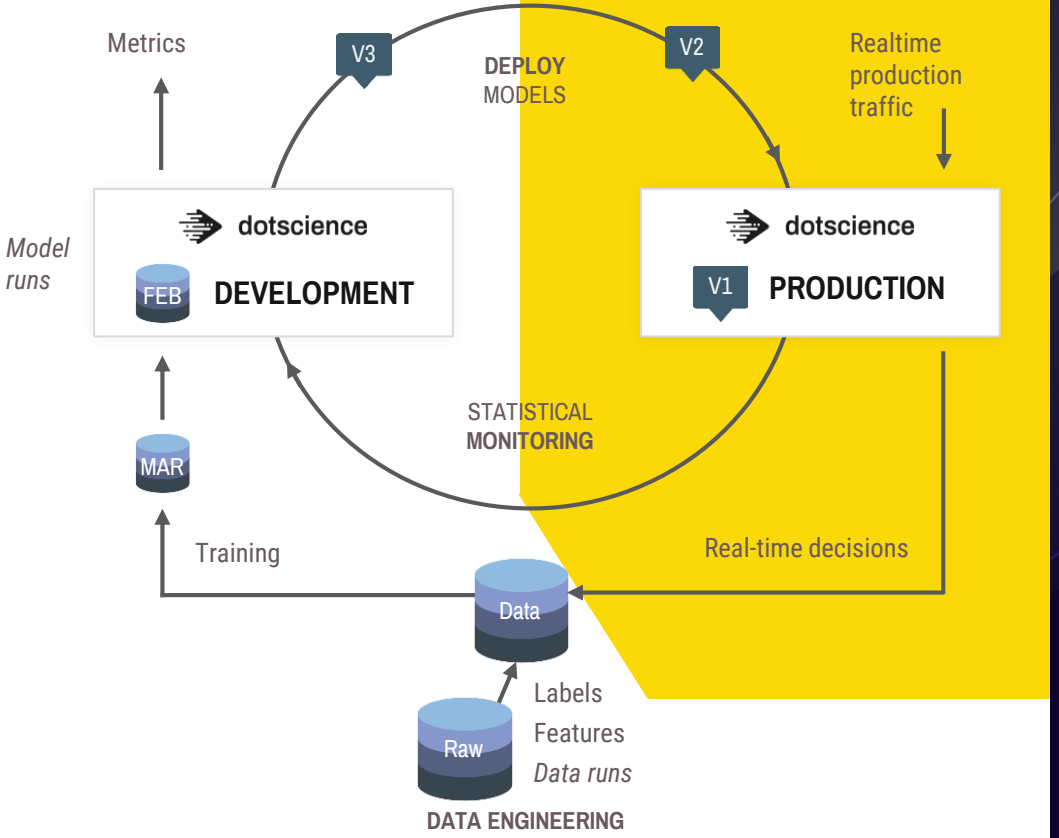
As raw data is processed, features engineered & samples annotated with labels, every data version needs to be recorded and made available for model development with full provenance. Need to avoid questions about which data was used to train a model.



Model Development

Once your data is annotated and ready to start building models, need to track model runs.

You can increase team productivity by building a 'runnable ML knowledge base' to eliminate silos. Need to reduce key person risk by making it easy for anyone to pick up where another left off.



Models in Production

Need to be able to push your models to production through a CI/CD system.

Once models are in production, need to keep them performing reliably. Need alerting on issues with statistical monitoring & to be confident changes to models to fix issues are working need ability to do forensic provenance tracking.

Dotscience key features

DEMO

Track runs in Jupyter & scripts

dotscience lab

← Resources

Activity →

File Edit View Run Kernel Tabs Settings Help

Dotscience

Name	Last Modified
model	4 minutes ago
get-data.ipynb	3 minutes ago
roadsigns.ipynb	3 minutes ago
signnames.csv	6 minutes ago
test.p	3 years ago
train.p	2 years ago
valid.p	2 years ago

get-data.ipynb | roadsigns.ipynb

Code | Python 3

```
[*]: num_classes=43
conv = tf.keras.models.Sequential()
conv.add(tf.keras.layers.Conv2D(32, kernel_size=(5, 5), strides=(1, 1), activation='relu', input_shape=(32, 3
conv.add(tf.keras.layers.MaxPooling2D(pool_size=(2, 2)))
conv.add(tf.keras.layers.Conv2D(32, (5, 5), activation='relu'))
conv.add(tf.keras.layers.MaxPooling2D(pool_size=(2, 2)))
conv.add(tf.keras.layers.Flatten())
conv.add(tf.keras.layers.Dense(1000, activation='relu'))
conv.add(tf.keras.layers.Dense(units = num_classes, activation='softmax'))

conv.compile(optimizer=ds.parameter("optimizer", 'adam'),
             loss='sparse_categorical_crossentropy',
             metrics=['accuracy'])

es = tf.keras.callbacks.EarlyStopping(monitor='acc')

conv.fit(X_train_preprocessed, y_train,
         epochs=ds.parameter("epochs", 3),
         verbose=1,
         validation_data=(X_valid_preprocessed, y_valid),
         callbacks=[es])
```

```
Train on 34799 samples, validate on 4410 samples
Epoch 1/3
19168/34799 [===============>..............] - ETA: 12s - loss: 0.9702 - acc: 0.7312
```

```
[*]: ds.summary("accuracy", conv.evaluate(X_test_preprocessed, y_test)[1])
```

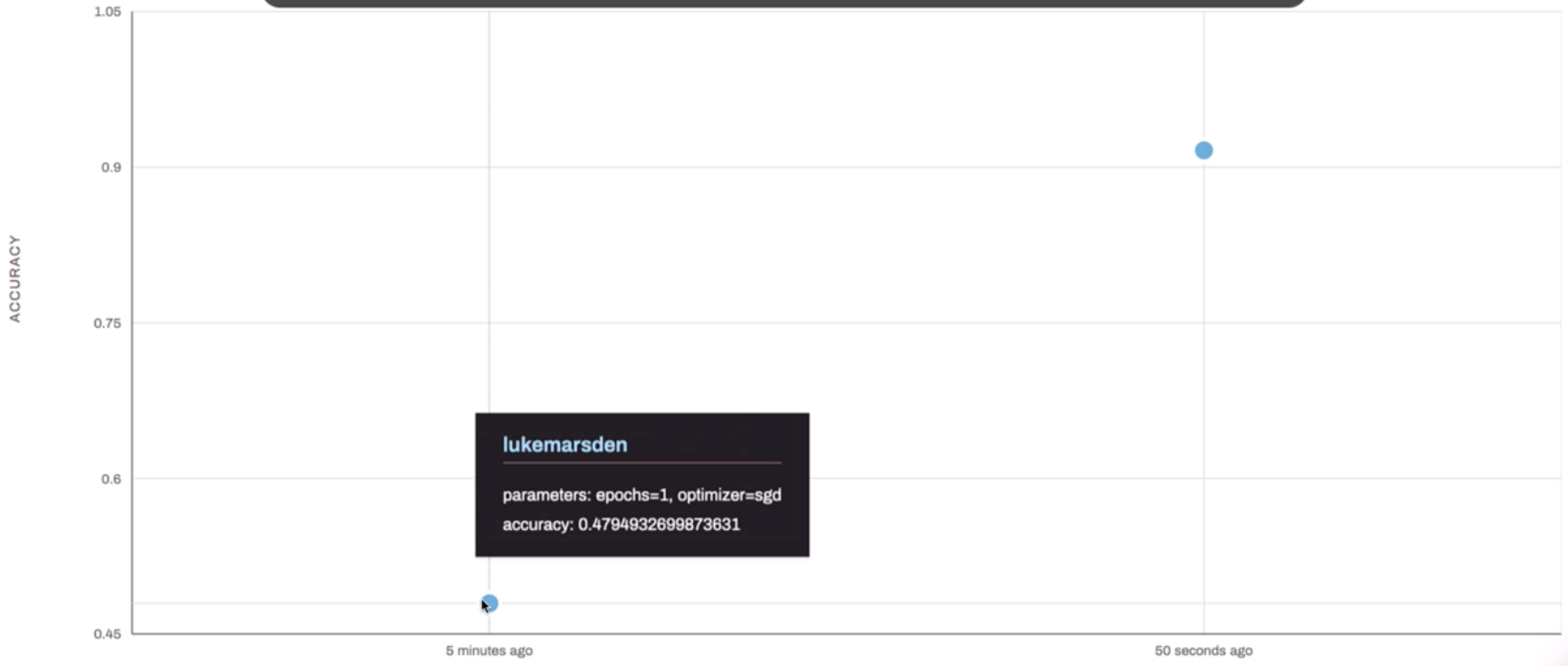
DEMO

dotscience

https://modeldemo.dotscience.net/projects/5d719c60-7709-48f9-ab02-514befe26bb5/activity/?direction=right

Plot chart

Record & share run metrics

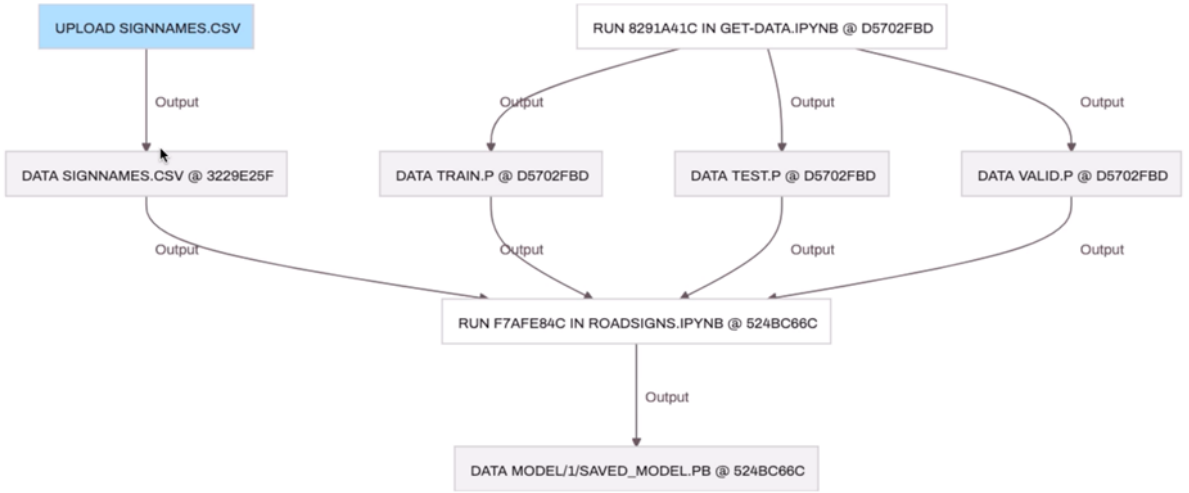


DEMO

Track provenance of data & models

Provenance graph

Reset zoom



DEMO

Deploy & monitor models

dotscience

Projects Models

Models

Your models

PROJECT	DATE/TIME	PARAMETERS	SUMMARY	FRAMEWORK	MODEL	DEPLOY
ROAD. SIGNS. YEAH.	7 May 2019 @ 15:28:34	epochs=1 optimizer=adam	accuracy=0.8938242280473807	tensorflow	./model	Deploy →
ROAD. SIGNS. YEAH.	7 May 2019 @ 15:28:34	epochs=1 optimizer=adam	accuracy=0.8938242280473807	tensorflow	./model	Deploy →
ROAD. SIGNS. YEAH.	7 May 2019 @ 20:05:10	epochs=1 optimizer=sgd	accuracy=0.5511480601269956	tensorflow	./model	Deploy →
ROAD. SIGNS. YEAH.	8 May 2019 @ 09:28:56	epochs=1 optimizer=rmsprop	accuracy=0.9194774346982121	tensorflow	./model	Deploy →
ROAD. SIGNS. YEAH.	8 May 2019 @ 09:28:56	epochs=2 optimizer=rmsprop	accuracy=0.9272367379444512	tensorflow	./model	Deploy →
ROAD. SIGNS. YEAH.	8 May 2019 @ 09:28:56	epochs=1 optimizer=rmsprop	accuracy=0.9155977830184611	tensorflow	./model	Deploy →
ROAD. SIGNS. YEAH.	8 May 2019 @ 09:28:56	epochs=2 optimizer=rmsprop	accuracy=0.9183689627775765	tensorflow	./model	Deploy →
ROAD. SIGNS. YEAH.	8 May 2019 @ 09:28:56	epochs=2 optimizer=adam	accuracy=0.9240696753949659	tensorflow	./model	Deploy →
ROAD. SIGNS. YEAH.	8 May 2019 @ 09:28:56	epochs=1 optimizer=sgd	accuracy=0.5727632620980224	tensorflow	./model	Deploy →

dotscience model demo

Road Sign Predictor

Click on an image below for the production deployed model to predict what it is...

Speed limit (60km/h)

Predict

No entry

Predict

Pedestrians

Predict

Stop

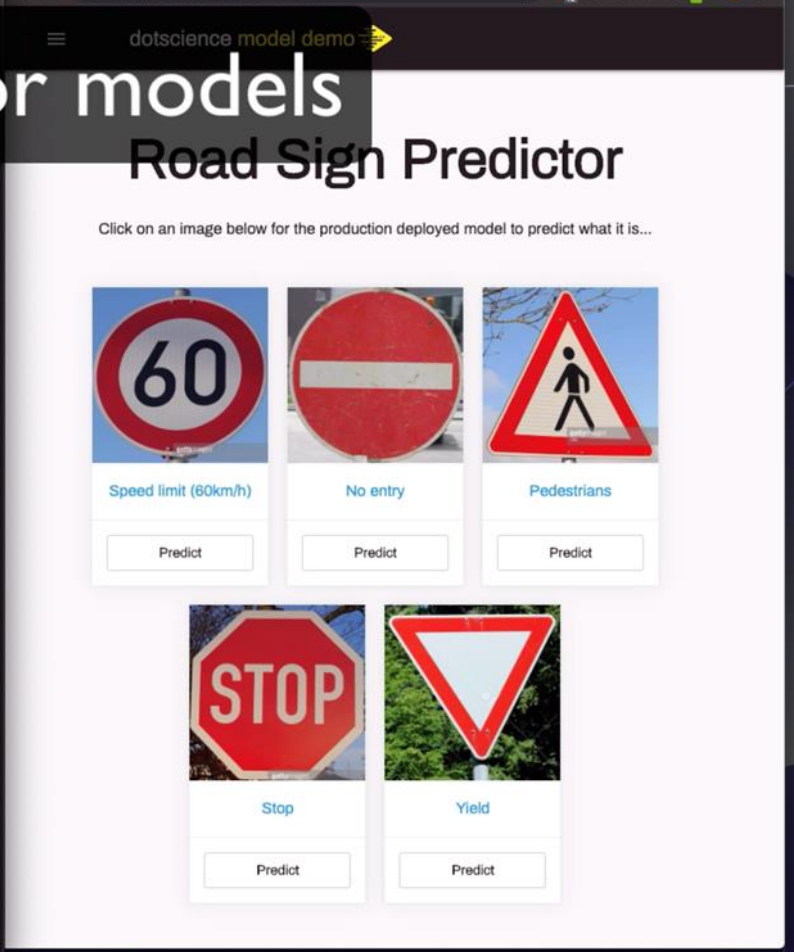
Predict

Yield

Predict

DEMO

Deploy & monitor models



TrueLayer are using Dotscience in Fintech

Luca Palmieri
Machine Learning and Data Engineering at TrueLayer

"The world of ML has a lot to learn from all the best practices developed to handle the Software Engineering lifecycle in the last 10 years. Dotscience has the potential to bring some of those hard-learned lessons to the ML world without forcing data scientists and researchers to completely abandon their tools of choice, like Jupyter Notebooks. It's a bold proposition and has the potential to make a huge impact."

TRUELAYER

Dotscience in the CORaiL project

Anders Åström
Datascience Manager at a global Technology Consulting firm

"The processes and tools for collaborating and maintaining ML projects at industrial scale is not yet as mature as for traditional software projects. The ML workflows pose several additional challenges that doesn't perfectly fit into Software DevOps processes. I am excited to work with Dotscience to tackle these challenges in our upcoming project, as they are actively focused on making collaboration structured and centralised so that it scales to much larger team and project sizes."

PROJECT CORaiL

Dotscience achieves the DevOps for ML requirements

1. Reproducible
Every run is fully reproducible. Code, data, parameters, metrics: it's all there, every detail.

2. Accountable
Full provenance of every model gives a complete forensic audit trail of where models came from.

3. Collaborative
Runs continuously published for team learning. Users can fork others' projects and make pull requests, notebooks included.

4. Continuous
Models can be deployed to production from Dotscience, and then statistically monitored.

Dotscience is a Highly Differentiated AI Platform

Accelerate AI projects
Optimal team productivity from run tracking across datasets, code, notebooks, models & metrics.

Strong model accountability
Ability to track back from any decision made in production guarantees compliance.

Runs anywhere
Zero dependencies so Dotscience runs anywhere, from laptop, to GPU rig, enterprise data center to cloud.

End-to-end AI Platform
From development to production, track full evolution of a model & metrics throughout the lifecycle.