How GIS and Machine Learning Work Together

Shannon Kalisky & Atma Mani

SEE
WHAT
OTHERS
CAN'T

Agenda

- What is machine learning?
- Data engineering
- Analysis and algorithms
- Industry use cases
- What is spatial machine learning?

What is machine learning outside of GIS?

My favorite definition:

"Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions."

- Emerj Artificial Intelligence Research

How can you have a favorite definition? Isn't it established?

- **"Machine Learning at its most basic is the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world."** – Nvidia
- **"Machine learning is the science of getting computers to act without being explicitly programmed."** – Stanford
- **"Machine learning is based on algorithms that can learn from data without relying on rules-based programming."**- McKinsey & Co.
- **"Machine learning algorithms can figure out how to perform important tasks by generalizing from examples."** – University of Washington
- **"The field of Machine Learning seeks to answer the question "How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?"** – Carnegie Mellon University

What is machine learning outside of GIS?

My favorite definition:

"Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions."

- Emerj Artificial Intelligence Research

GIS

People

Deliverables

Notebooks

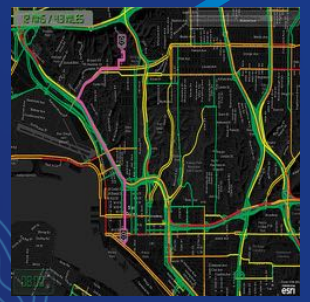
Apps

Maps

Data

ArcGIS

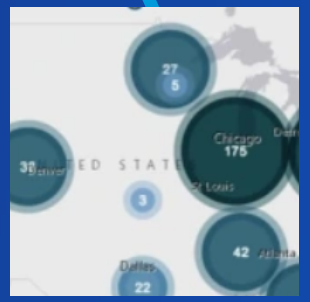
Data



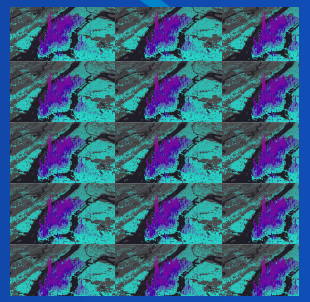
Real-time

Imagery/Raster

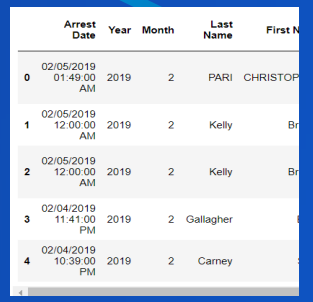
Vector



BI



Big data



	Arrest Date	Year	Month	Last Name	First N
0	02/05/2019 01:49:00 AM	2019	2	PARI	CHRISTOP
1	02/05/2019 12:00:00 AM	2019	2	Kelly	Br
2	02/05/2019 12:00:00 AM	2019	2	Kelly	Br
3	02/04/2019 11:41:00 PM	2019	2	Gallagher	
4	02/04/2019 10:39:00 PM	2019	2	Carney	

Tabular

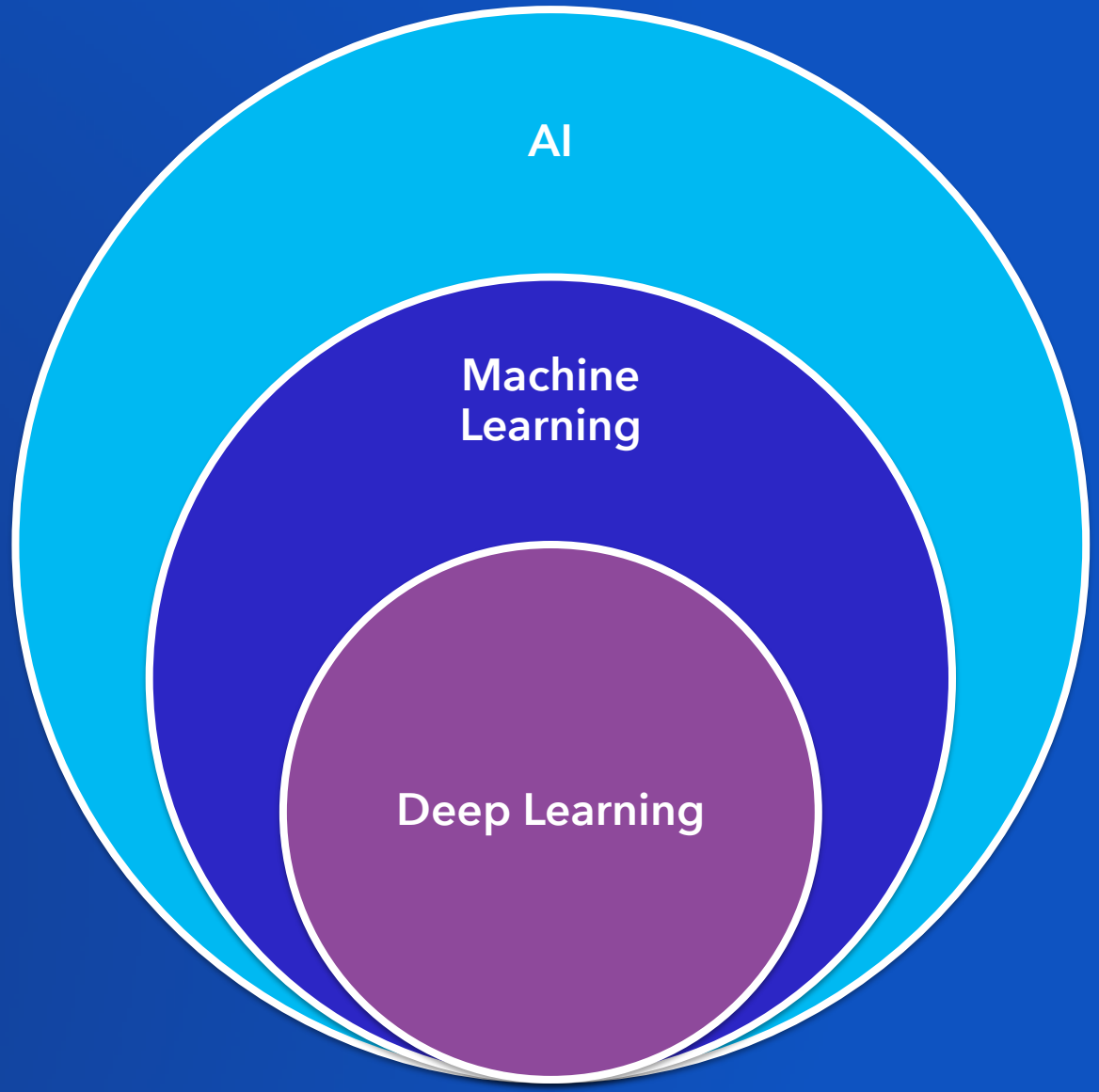
Machine Learning within ArcGIS

Click-button interface:

- **ArcGIS Pro**
  - **geoprocessing tools**
  - **ModelBuilder**
- **GeoAnalytics, web tools**

Scripting interface:

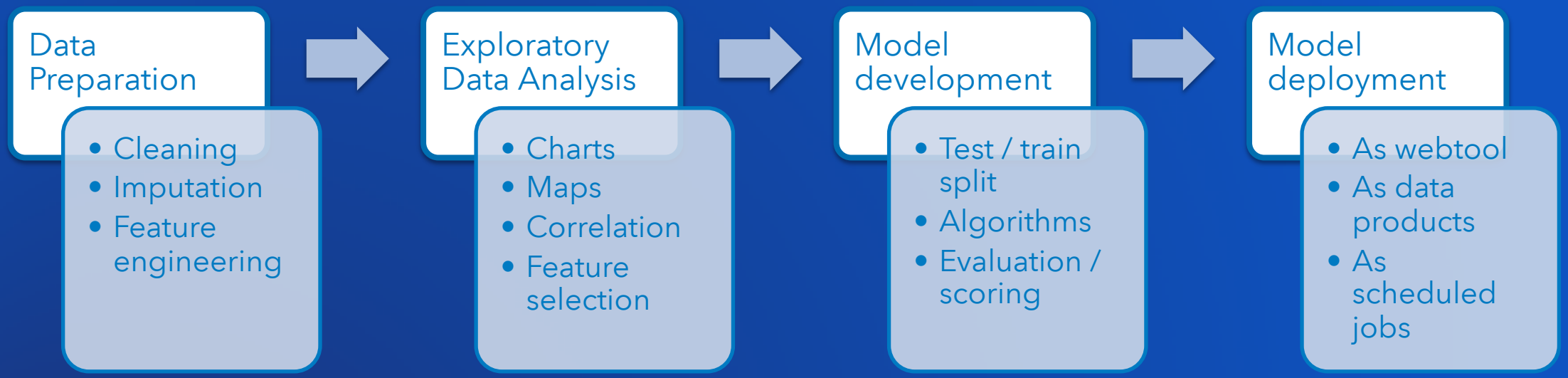
- **ArcPy**
- **ArcGIS API for Python**
  - **arcgis.learn module (deep learning)**
- **ArcGIS Notebooks**
- **Open source Python libraries**
  - **scikit-learn, TensorFlow, CNTK, Keras, …**



How do GIS and Machine Learning work together?

1 Machine learning needs information and data from observations and real-world interactions. GIS is an authoritative system of record for such data.

Machine learning methodology



Data Engineering

In the non-GIS data science world…

Features

```
In [9]: import pandas as pd

In [10]: df = pd.read_csv(arrests)
         df.head()
```

Out[10]:

	Arrest Date	Year	Month	Last Name	First Name	Gender	Race	Ethnicity	Year of Birth	Age	From Address	From City	From State	Statute Type	Statute Code	Statute Des
0	02/05/2019 01:49:00 AM	2019	2	PARI	CHRISTOPHER	Male	White	Non-Hispanic	1964	54	157 WOOSTER AVE	RIVERSIDE	Rhode Island	RI Statute Violation	21-28-4.01-C1A	POSSESSIO O SCHEDULE II I
1	02/05/2019 12:00:00 AM	2019	2	Kelly	Bridget	Female	White	Non-Hispanic	1969	29	5 NATHANIEL WAY	canton	NaN	RI Statute Violation	31-27-2.1	Chemical Te Refus
2	02/05/2019 12:00:00 AM	2019	2	Kelly	Bridget	Female	White	Non-Hispanic	1969	29	5 NATHANIEL WAY	canton	NaN	RI Statute Violation	31-27-2	Driving Unde the Influenc of Liquor c Drugs
3	02/04/2019 11:41:00 PM	2019	2	Gallagher	Brien	Male	White	Non-Hispanic	1980	38	2 DIXON ST	Providence	NaN	RI Statute Violation	11-45-1	DISORDERL CONDUC
4	02/04/2019 10:39:00 PM	2019	2	Carney	Scott	Male	White	Non-Hispanic	1987	31	153 EDGEHILL RD	Norwood	NaN	RI Statute Violation	11-4-6	ARSON-5TI DEGRE

In GIS world…

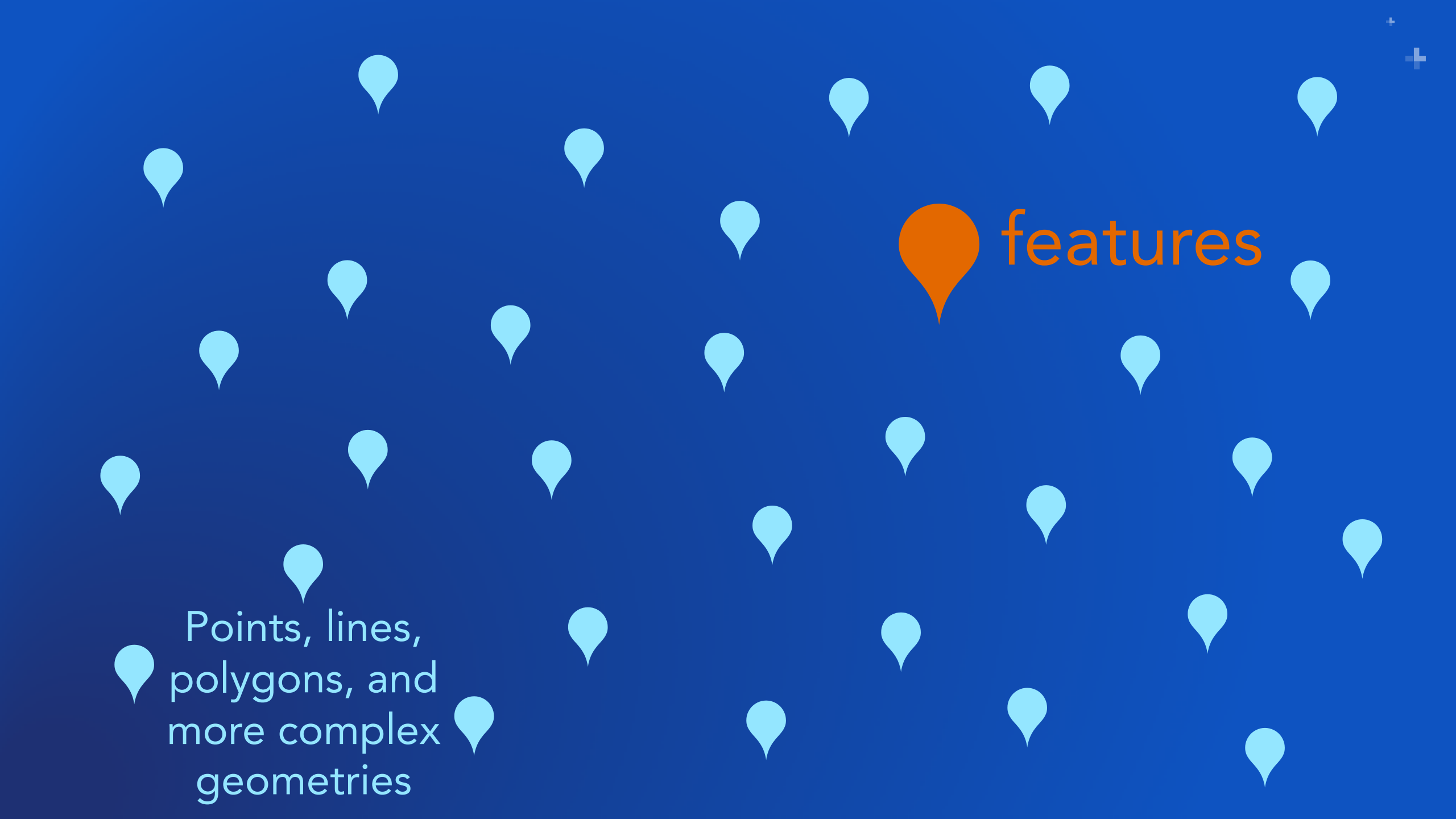
Attributes

```
In [9]: import pandas as pd

In [10]: df = pd.read_csv(arrests)
         df.head()
```

Out[10]:

	Arrest Date	Year	Month	Last Name	First Name	Gender	Race	Ethnicity	Year of Birth	Age	From Address	From City	From State	Statute Type	Statute Code	Statute Des
0	02/05/2019 01:49:00 AM	2019	2	PARI	CHRISTOPHER	Male	White	Non-Hispanic	1964	54	157 WOOSTER AVE	RIVERSIDE	Rhode Island	RI Statute Violation	21-28-4.01-C1A	POSSESSIO O SCHEDULE II I
1	02/05/2019 12:00:00 AM	2019	2	Kelly	Bridget	Female	White	Non-Hispanic	1969	29	5 NATHANIEL WAY	canton	NaN	RI Statute Violation	31-27-2.1	Chemical Tes Refusa
2	02/05/2019 12:00:00 AM	2019	2	Kelly	Bridget	Female	White	Non-Hispanic	1969	29	5 NATHANIEL WAY	canton	NaN	RI Statute Violation	31-27-2	Driving Unde the Influenc of Liquor c Drugs
3	02/04/2019 11:41:00 PM	2019	2	Gallagher	Brien	Male	White	Non-Hispanic	1980	38	2 DIXON ST	Providence	NaN	RI Statute Violation	11-45-1	DISORDERL CONDUC
4	02/04/2019 10:39:00 PM	2019	2	Carney	Scott	Male	White	Non-Hispanic	1987	31	153 EDGEHILL RD	Norwood	NaN	RI Statute Violation	11-4-6	ARSON-5TI DEGRE


features
Points, lines, polygons, and more complex geometries

Feature engineering

- Feature engineering is a concept in machine learning
  - Remember, features in machine learning = attributes in GIS
- Feature engineering is basically data cleaning + data preparation so that bad or noisy data is eliminated or minimized in the analysis process
  - Sometimes data cleaning and feature engineering are viewed as two different tasks

Common feature engineering methods

- **Imputation → dealing with missing values**
  - Dropping rows or columns that have missing data – not the most ideal
  - Numerical imputation – filling missing data with a number value (such as 0 or the median value of the column)
  - Categorical imputation – filling with a value such as "Other"
- **Handling outliers**
  - Detect outliers with standard deviation
  - Detect outliers with percentiles
  - Drop the row or cap the value?
- **Binning**
- **Logarithmic transformation (log transform)**
  - Good for handling skewed data
  - Data must be numeric and have all positive values
- **Create dummy values**
  - One-hot encoding: Encoding data as 0 or 1
- **Scaling**
  - Normalization
  - Standardization

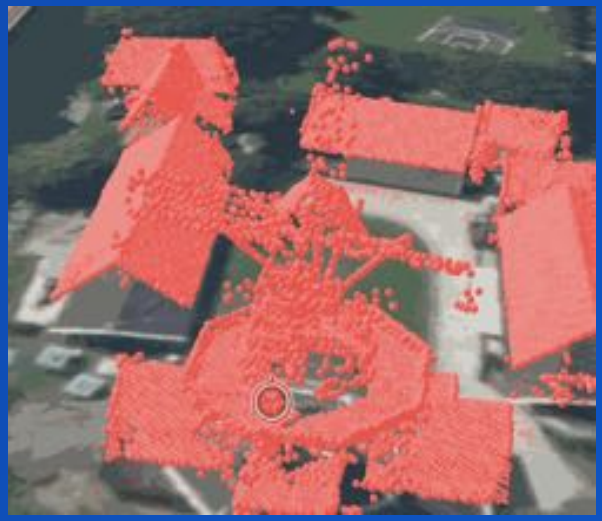
Spatial feature engineering

- Geoenrichment
- Overlay analysis
- Find outliers
- Space-time binning
- Imputation using nearest neighbors
- Engineering data based on geography
  - Distance, area, drive time, elevation, etc.

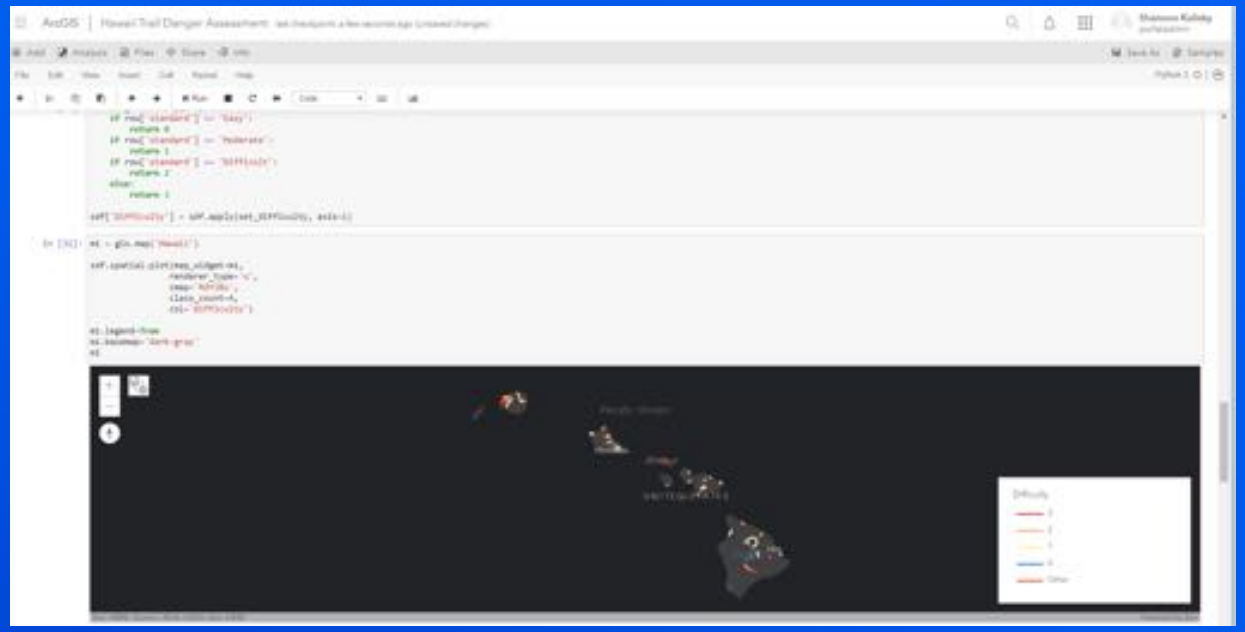
Data exploration

- Machine learning does not introduce new data exploration methods, rather existing data exploration methods apply.
  - Charts
  - Graphs
  - Data tables (pandas dataframes)

- GIS adds spatiotemporal data exploration methods for tabular, 2D, 3D, imagery, and raster spatial exploration.



- ArcGIS brings machine learning data exploration and spatial exploration methods together



Data Engineering and Exploration

Demo

How do GIS and Machine Learning work together?

2

GIS provides more options for feature engineering and data exploration that may increase the accuracy and depth of the machine learning model.

Analysis and algorithms

Popular machine learning algorithms

- **Non spatial**
  - Generalized Linear Models
    - Linear regression, Lasso, Ridge, Logistic
  - Random forests Classification, Regression
  - Support Vector Classification, Regression
  - K-means clustering
  - Deep learning

- **Spatial**
  - Generalized Linear Models
  - Geographically Weighted Regression
  - Empirical Bayesian Kriging / Regression
  - Imagery classification algorithms
    - Random Forest Classifier
    - Support Vector Classifier
    - Maximum Likelihood Classifier
    - ISO classifier
  - Clustering
    - Density based clustering
    - Optimized hot spot analysis

Popular machine learning libraries

- **Spatial**
  - ArcPy
  - ArcGIS API for Python
  - PySAL
- **Non Spatial**
  - Scikit-learn
  - Scikit-image
  - Xgboost
  - Adaboost
  - PyTorch, Fast.ai
  - Tensorflow, Keras
  - CNTK, Caffe

Machine learning jargon

Demystifying machine learning jargon – Data Engineering

Term	Meaning
Dimensionality reduction	Reducing number of attributes / columns / bands. Improves the model when highly correlated columns are removed. Methods: PCA
Normalization / transformation	Scale attribute values to fall within a known range. Improves the model when columns have widely varying min-max, units. Methods: Gaussian, Min-Max Scaler, Logarithmic, Fourier, Coordinate transformations
Imputation	Fill for missing values. Retains size of training data instead of tossing out empty rows. Methods: Median, Mean, Spatial imputation
Enrichment	Add relevant columns / attributes / bands. Improves the model as it can describe the phenomena better. Examples: Weather, Geoenrichment, Stock index, etc.

Demystifying machine learning jargon – Exploratory Data Analysis

Term	Meaning
Clustering	Find natural groupings in data. Methods: K-means, mea-shift, DB Scan, HDBScan, hot spot analysis
Feature importance	Sort / discover attributes by their influence or importance. Methods: Regression, Tree based models
Hypothesis testing	Validate business assumptions. Example: Are women more likely to use this app? Does distance to highway influence # of customer visits in a store?. Methods: T-test, ANOVA

Demystifying machine learning jargon – Models & Algorithms - General

Term	Meaning
Supervised ML	Teach a model using existing training data. Predict on unknown
Unsupervised ML	Algorithms which figure out relationships and facts from data without human intervention
Overfitting	When models get too complex and starts modeling noises in training data.
Underfitting	When model inadequately characterizes variability in data
Bias variance trade off	Bias – error in prediction. Variance – how robust model remains for slight modifications in training data.
Naïve Bayes	Prediction based on Bayes theorem. Formulates probability of output (prediction) based on probabilities of features.
Classification	Random Forests, SVM, XGBoost, ADAboost
Monte Carlo simulation	Technique used to understand risk in forecasting models. Presents ranges of outcomes and their probabilities.

Demystifying machine learning jargon – Models & Algorithms – Deep Learning

Term	Meaning
Deep Neural Networks (DNNs)	A type of Artificial Neural Nets that have several hidden layers. Capable of higher level abstraction from large, unstructured, diverse input data.
Stochastic Gradient descent	Method by which a DNN updates the weights of the layers in the model. This is an optimization technique by which algorithm knows which direction to update the weights (increase / decrease / no change)
Learning rate	Specifies the amount by which weights are updated while learning. Influences how much new data will override learnt knowledge.
Epochs	The number of passes a DNN takes on the training samples.
ResNET	Residual Network, a type of Neural Net architecture. Other examples: LeNET, AlexNet, GoogLeNet

What is spatial machine learning?

Spatial machine learning is when spatial data, methods, or algorithms are applied to the machine learning process.

So…

Spatial Machine Learning *is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion,* ***factoring in the spatiotemporal context of data****, by feeding them* ***spatial data and location information*** *in the form of observations and real-world interactions.*

Spatial Data Science

Data Engineering and Exploration

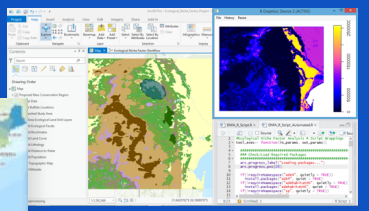
Spatial tools and algorithms

- Generalized Linear Regression
- Logistic GWR
- Forest-based Classification & Regression
- 3D EBK Regression
- EBK Regression Prediction
- Density-based Clustering
- Training data preparation
- Spatial optimization
- ArcPy
- arcgis.learn

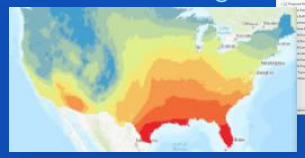
Data Science libraries and frameworks

- Conda
- Jupyter
- scikit-learn
- TensorFlow
- PyTorch
- Keras
- CNTK
- IBM Watson
- R
- Open source Python libraries

Conservation

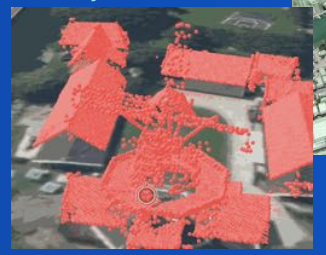


Climate Modeling



Reconstruct 3D Buildings

Classify Point Clouds

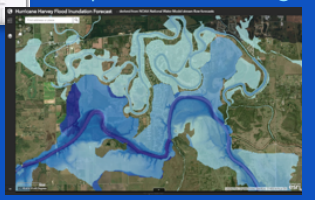


Landcover Classification

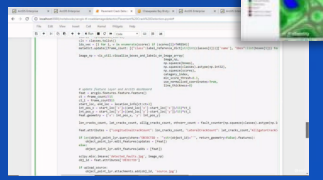
Accident Prediction

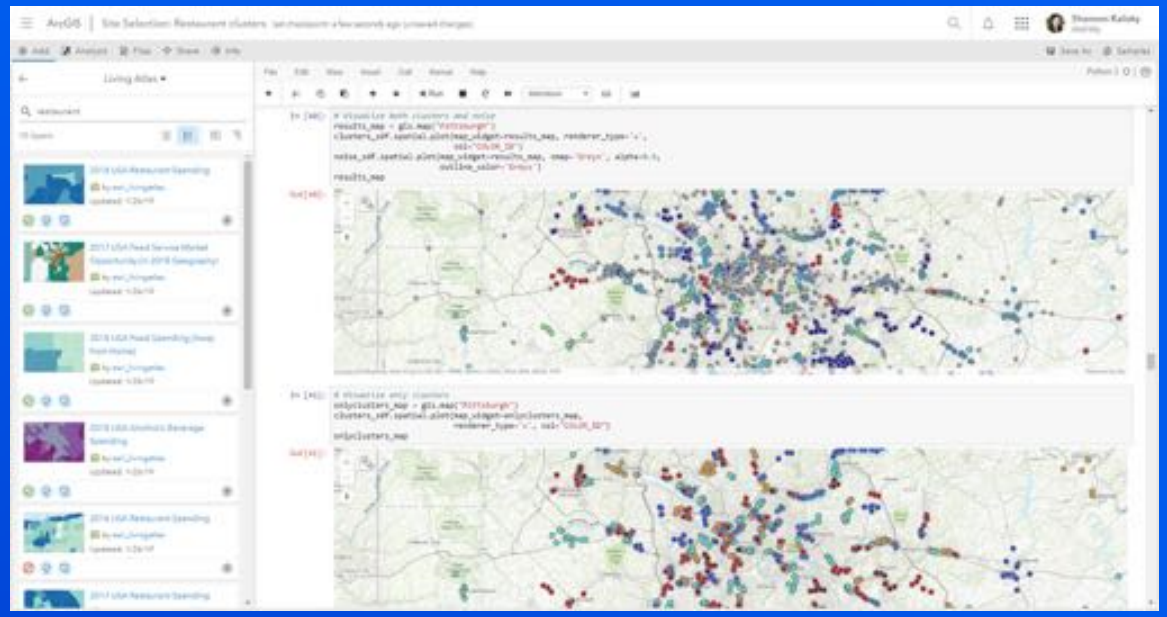
Feature Extraction

Impact Forecasting



Feature Identification





Spatial Machine Learning

Demo

Dive deeper into spatial machine learning methods:

- **Session:**
- **Machine Learning in ArcGIS**
  - Wednesday, July 10
  - 1:00 pm - 2:00 pm
  - SDCC - Room 10

  - Thursday, July 11
  - 1:00 pm - 2:00 pm
  - SDCC - Room 01 A/B

How do GIS and Machine Learning work together?

3 The spatial methods and algorithms of GIS can uncover patterns, outliers, and anomalies missed in non-spatial machine learning.

What do people do with machine learning?

Section Subhead

The purpose of machine learning is to solve problems.

Machine learning can be used in cases where using human resources is not time/cost effective or when many variables are being considered simultaneously.

Examples of machine learning and AI across industries

- **Electric, Gas, and Water Utilities**
  - Predictive maintenance
  - Customer segmentation
  - Impact predictions
- **Petroleum and Pipeline**
  - Predictive maintenance
  - Human & environmental safety
  - High yield opportunities
- **Natural Resources**
  - Impact predictions
  - Precision agriculture
  - Human & environmental safety
- **Transportation**
  - Route optimization
  - Predictive maintenance
  - Human & environmental safety
- **Manufacturing**
  - Predictive maintenance
  - Logistics
  - Anomaly detection
- **Insurance**
  - Customer segmentation
  - Fraud detection
  - Claims prediction
- **Banking**
  - Customer segmentation
  - Risk modeling
  - Site selection
- **Retail**
  - Customer segmentation
  - Logistics
  - Anomaly detection
- **Real Estate**
  - Site selection
  - Value predictions
  - Target marketing
- **Telecommunications**
  - Predictive maintenance
  - Fraud detection
  - Target marketing
- **Health and Human Services**
  - Customer segmentation
  - Logistics
  - Research
- **Government**
  - Fraud detection
  - Research
  - Impact assessments
- **State and Local Government**
  - Route optimization
  - Sites selection
  - Enable citizen data scientists
- **Public Safety**
  - Route optimization
  - Predictive maintenance
  - Impact assessment
- **Sustainable Development**
  - Route optimization
  - Site selection
  - Research

…and more

Esri examples:

- Climate modeling
- Impact/damage assessments
- Accident prediction
- Sales forecasting
- Predictive site selection
- Estimated time of arrival prediction
- Smart road digitization
- Feature extraction from LiDAR point clouds
- Identifying urban heat islands
- Identifying and classifying pavement cracks
- Object detection, pixel classification, and image segmentation with deep learning
- …

Dive deeper into use case examples:

Session:

GeoAI: Vertical Use-cases for applying AI with ArcGIS

- Wednesday, July 10
- 10:00 am - 11:00 am
- SDCC - Room 14 B

How do GIS and Machine Learning work together?

4 Many of the problems organizations wish to solve with machine learning and AI are spatial problems and can benefit from the use of spatial data and methods in the data engineering, analysis, and model building process.

Machine learning, AI, and Ethics

Section Subhead

Explainable AI > An AI “black box”

- **AI itself is a black box**
  - We understand the design of the model, but can’t say how exactly it made a prediction
- **Privacy and business liability in the age of algorithms**
  - Still being debated around the world, the next few years will be transformational
  - AI algorithms are hungry creatures – GIS Analysts are gatekeepers of organizational data
- **We use AI to drive action, our actions impact the world**
  - GIS is the software that is designed to quantify the impact that human actions have on the human and natural worlds
    - **GIS is more important than ever.**

What does this mean for machine learning and GIS?

- Understand the data that goes into your model
  - Avoid overfitting
- Aim for reproducibility and transparency
  - In notebooks:
    - Clear all cell output, adjust, and run all → reproduce your own work before you say it's ready
    - Use markdown to explain analysis steps and logic
    - Use the Web GIS to have explanatory item details pages
    - Add web maps, apps, and rich media to add context and story to your notebook
- Modularize workflows
- Make sure all data needed for the model is shared and documented
- Do not set it and forget it
  - Revisit analyses and models to make sure that the model still fits the data

How do GIS and Machine Learning work together?

5 AI is naturally a black box, but using GIS to understand data and quantify the impact of algorithm driven analyses can help create ethical and explainable AI.

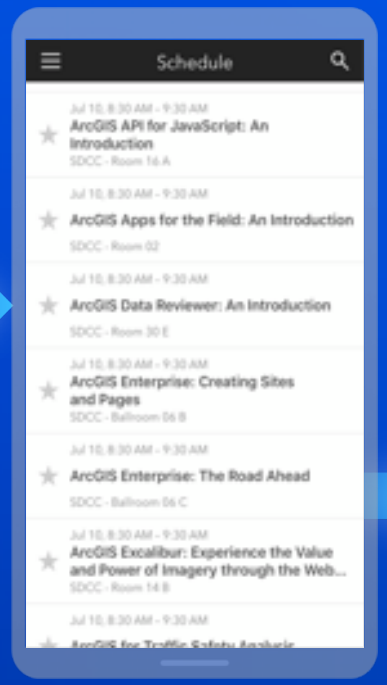
Summary: How do GIS and Machine Learning work together?

1) **Access to data**
2) **More options for data engineering and exploration**
3) **Find patterns, outliers, and anomalies missed by non-spatial methods**
4) **Many machine learning problems are spatial**
5) **GIS can help drive explainable and ethical AI**

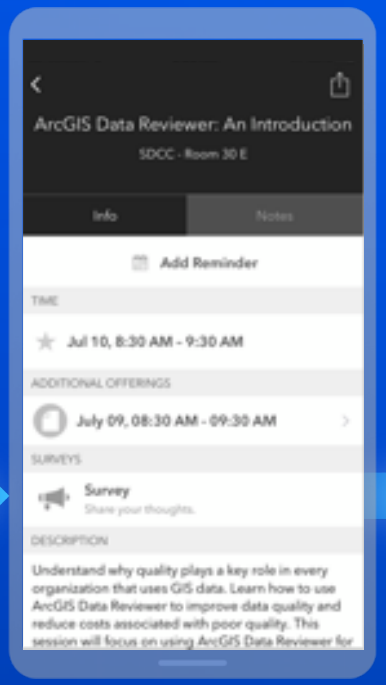
Please Share Your Feedback in the App

Download the Esri Events app and find your event

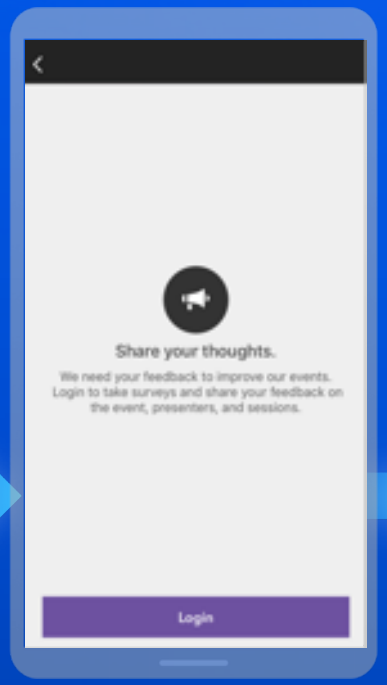
Select the session you attended



Scroll down to "Survey"



Log in to access the survey



Complete the survey and select "Submit"