CSCE 590-1: From Data to Decisions with Open Data: A Practical Introduction to AI

Lecture 12: Advanced Machine Learning Topics

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

18TH FEB, 2021

Carolinian Creed: "I will practice personal and academic integrity."

Organization of Lecture 12

- Introduction Segment
  - Recap of Lecture 11
- Main Segment
  - Generating Explanations
    - LIME
  - AutoAI
- Concluding Segment
  - Quiz 2
  - About Next Lecture – Lecture 13
  - Ask me anything

Introduction Segment

Recap of Lecture 11

- Clustering methods
- Distance metrics
- Measuring cluster quality
- Explaining / describing clusters

Main Segment

Generating Explanations

AI Explainability

Meaningful explanations depend on the explanation consumer

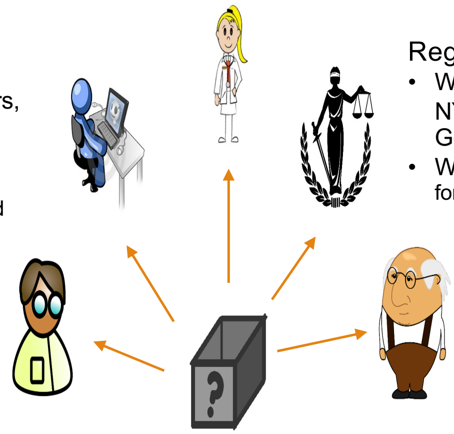
he General Data Protection Regulation (GDPR)

- Limits to decision-making based solely on automated processing an profiling (Art.22)
- Right to be provided with meaningful information about the logic involved in the decision (Art.13 (2) f. and 15 (1) h)

End Users
- Who: Physicians, judges, loan officers, teacher evaluators
- Why: trust/confidence, insights(

AI System builders, stakeholders
- Who: data scientists, developers, prod mgrs
- Why: ensure/improve performance

Regulatory Bodies
- Who: EU (GDPR), NYC Council, US Gov't, etc
- Why: ensure fairness for constituents

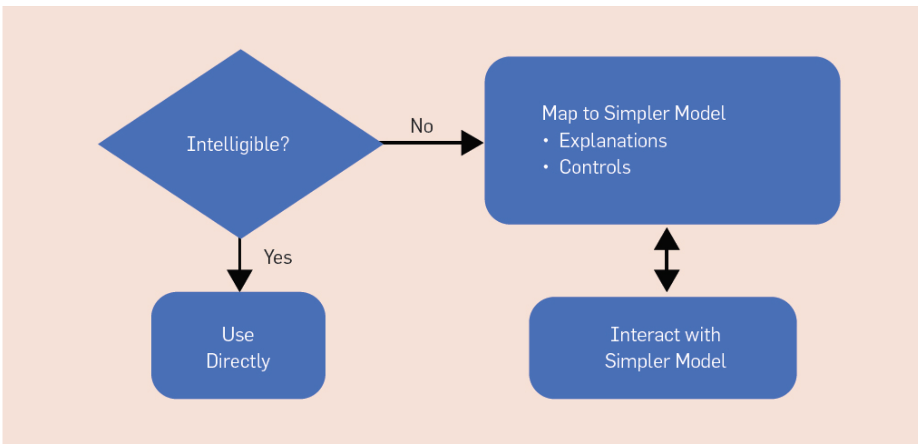


Affected Users
- Who: Patients, accused, loan applicants, teachers
- Why: understanding of factors

Must match the complexity capability of the consumer

Must match the domain knowledge of the consumer

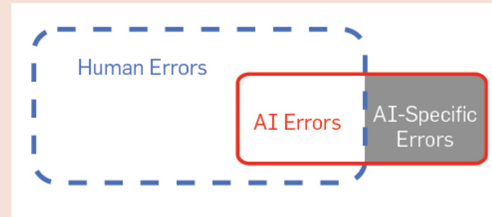
Setting and Terminology: Intelligible Models and Explanations



- Transparency: providing stakeholders with relevant information about how a model works
- Explainability: Providing insights into model's behavior for specific datapoints

Sources:
1. The Challenge of Crafting Intelligible Intelligence, Daniel S. Weld, Gagan Bansal, Communications of the ACM, June 2019, Vol. 62 No. 6, Pages 70-79, 10.1145/3282486
2. Explainable Machine Learning in Deployment, FAT* 2020.

Need for Intelligibility



The red shape denotes the AI's mistakes; its smaller size indicates a net reduction in the number of errors. The gray region denotes AI-specific mistakes a human would never make. Despite reducing the total number of errors, a deployed model may create new areas of liability (gray), necessitating explanations.

- **AI may have the wrong objective:** is AI right for the right reasons?
- **AI may be using inadequate features:** understand modeling issues
- **Distributional drift:** detect when and why models are failing to generalize
- **Facilitating user control:** guiding what preferences to learn
- **User acceptance:** especially for costly actions
- **Improving human insight:** improve algorithm design
- **Legal imperatives**

Source: The Challenge of Crafting Intelligible Intelligence, Daniel S. Weld, Gagan Bansal, Communications of the ACM, June 2019, Vol. 62 No. 6, Pages 70-79, 10.1145/3282486

Types of Explanations

- **Feature-based**: from the features of the data, which feature(s) were most important for given decision output
  - Example: For a loan, is it income or the person's age ?
- **Sample-based**: from data in training, which data points were important for given test point; helps understand sampling and its representation in wider population
  - Example: For a loan, what instances similar to the loan application would have gotten the loan ?
- **Counter-factual**: what-ifs – what do you change about the input to change the decision output
  - Example: For a loan, does getting an additional borrower insurance increase chance of getting the loan?
- Natural language

Source: Explainable Machine Learning in Deployment, FAT* 2020

Stakeholders for Explanations

- Executives
  - Explainability as a market differentiator. Do we need explanations?
- ML engineers
  - How to improve model's performance?
- End-users
  - Understand business decisions emanating from usage of AI
    - Why was my load denied?
    - Why a particular treatment was recommended or de-prioritized ?
- Regulators
  - Prove that you did not discriminate based on existing laws

Source: Explainable Machine Learning in Deployment, FAT* 2020

References for AI Explainability

Papers

- The Challenge of Crafting Intelligible Intelligence, Daniel S. Weld, Gagan Bansal, Communications of the ACM, June 2019, Vol. 62 No. 6, Pages 70-79, 10.1145/3282486
- "Why Should I Trust You?" Explaining the Predictions of Any Classifier, Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin, in ACM's Conference on Knowledge Discovery and Data Mining, KDD2016; https://homes.cs.washington.edu/~marcotcr/blog/lime/, https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/
- Explainable Machine Learning in Deployment, FAT* 2020, https://arxiv.org/pdf/1909.06342.pdf; Video: https://www.youtube.com/watch?v=Hofl4uwxtPA

Tutorial: XAI tutorial at AAAI 2020, https://xaitutorial2020.github.io/

Tool: AIX 360

Tool: https://aix360.mybluemix.net/

Video: https://www.youtube.com/watch?v=Yn4yduyoQh4

Paper: https://arxiv.org/abs/1909.03012

LIME – Local Interpretable Model-Agnostic Explanations

Paper: "Why Should I Trust You?" Explaining the Predictions of Any Classifier, Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin, ACM's Conference on Knowledge Discovery and Data Mining, KDD2016

Blogs:

- https://homes.cs.washington.edu/~marcotcr/blog/lime/
- https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/

Code: https://github.com/marcotcr/lime

LIME on Image

Question: Why is this a frog?

Divide image into interpretable components - contiguous superpixels

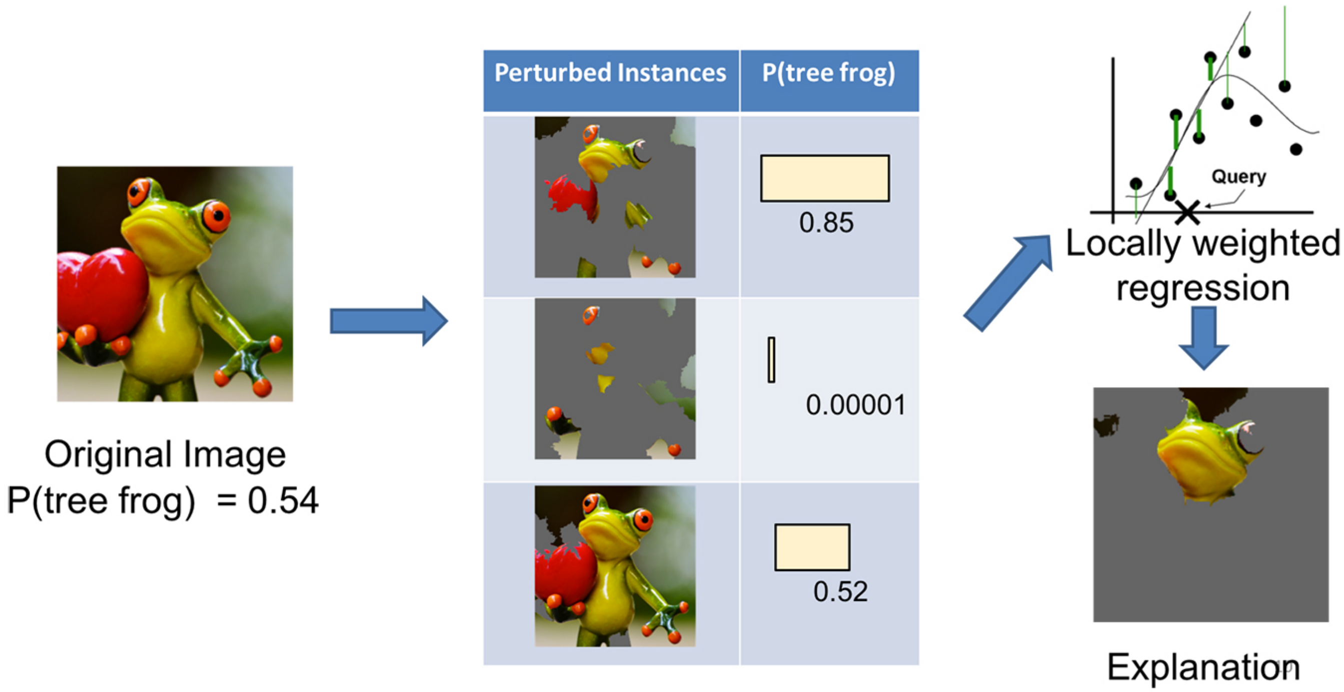
Original Image

Interpretable Components

Source: https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/

LIME

1. Generate a data set of perturbed instances by turning some of the interpretable components "off" (gray).
2. For each perturbed instance, calculate probability that a tree frog is in the image according to the model.
3. Learn a simple (linear) model on this data set, which is locally weighted
4. Output regions with highest positive weights as an explanation, graying out everything else.

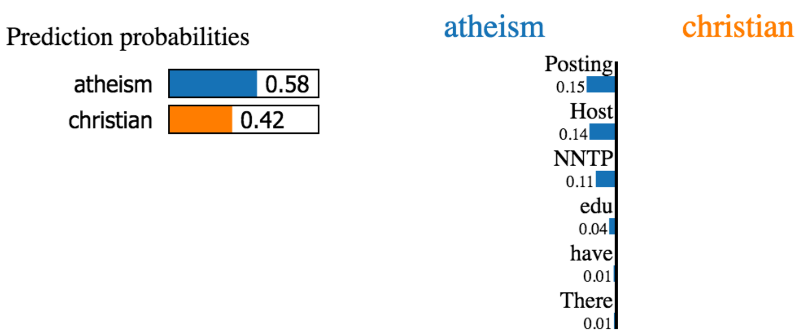


Source: https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/

LIME on Text

Question: Why is a classifier with >90% accuracy predicting based on ?

"iI we remove the words Host and NNTP from the document, we expect the classifier to predict atheism with probability 0.58 - 0.14 - 0.11 = 0.31".



Text with highlighted words

From: johnchad@triton.unm.edu (jchadwic)
Subject: Another request for Darwin Fish
Organization: University of New Mexico, Albuquerque
Lines: 11
NNTP-Posting-Host: triton.unm.edu

Hello Gang,

There have been some notes recently asking where to obtain the DARWIN fish.
This is the same question I have and I have not seen an answer on the
net. If anyone has a contact please post on the net or email me.

Source: https://github.com/marcotcr/lime

Code Examples

- Lime
  - https://github.com/biplav-s/course-d2d-ai/tree/main/sample-code/l12-explanability-autoai

- We will see:
  - Regression: https://github.com/biplav-s/course-d2d-ai/blob/main/sample-code/l12-explanability-autoai/explanation-lime.ipynb
  - Classification: https://github.com/biplav-s/course-d2d-ai/blob/main/sample-code/l12-explanability-autoai/explanation-lime-classification.ipynb

AIX 360

- Credit Approval Tutorial [on nbviewer]
Shows how to explain credit approval models that use the FICO Explainable Machine Learning Challenge dataset.
- Medical Expenditure Tutorial [on nbviewer]
Shows how to create interpretable machine learning models in a care management scenario using Medical Expenditure Panel Survey data.
- Dermoscopy [on nbviewer]
Shows how to explain dermoscopic image datasets used to train machine learning models that help physicians diagnose skin diseases.
- Health and Nutrition Survey [on nbviewer]
Shows how to quickly understand the National Health and Nutrition Examination Survey datasets to hasten research in epidemiology and health policy.
- Proactive Retention [on nbviewer]
Shows how to explain predictions of a model that recommends employees for retention actions from a synthesized human resources dataset.

FICO Explainable Machine Learning Challenge Data

- Anonymous dataset of Home Equity Line of Credit (HELOC) applications made by real homeowners

- The customers in this dataset have requested a credit line in the range of $5,000 - $150,000.

- The fundamental task is to use the information about the applicant in their credit report to predict whether they will make timely payments over a two-year period.

Dataset source: *https://community.fico.com/s/explainable-machine-learning-challenge?tabset-3158a=2*
Tutorial notebook: *https://github.com/IBM/AIX360/blob/master/examples/tutorials/HELOC.ipynb*

Field	Meaning	Monotonicity Constraint (with respect to probability of bad = 1)
ExternalRiskEstimate	Consolidated version of risk markers	Monotonically Decreasing
MSinceOldestTradeOpen	Months Since Oldest Trade Open	Monotonically Decreasing
MSinceMostRecentTradeOpen	Months Since Most Recent Trade Open	Monotonically Decreasing
AverageMInFile	Average Months in File	Monotonically Decreasing
NumSatisfactoryTrades	Number Satisfactory Trades	Monotonically Decreasing
NumTrades60Ever2DerogPubRec	Number Trades 60+ Ever	Monotonically Decreasing
NumTrades90Ever2DerogPubRec	Number Trades 90+ Ever	Monotonically Decreasing
PercentTradesNeverDelq	Percent Trades Never Delinquent	Monotonically Decreasing
MSinceMostRecentDelq	Months Since Most Recent Delinquency	Monotonically Decreasing
MaxDelq2PublicRecLast12M	Max Delq/Public Records Last 12 Months. See tab "MaxDelq" for each category	Values 0-7 are monotonically decreasing
MaxDelqEver	Max Delinquency Ever. See tab "MaxDelq" for each category	Values 2-8 are monotonically decreasing
NumTotalTrades	Number of Total Trades (total number of credit accounts)	No constraint
NumTradesOpeninLast12M	Number of Trades Open in Last 12 Months	Monotonically Increasing
PercentInstallTrades	Percent Installment Trades	No constraint
MSinceMostRecentInqexcl7days	Months Since Most Recent Inq excl 7days	Monotonically Decreasing
NumInqLast6M	Number of Inq Last 6 Months	Monotonically Increasing
NumInqLast6Mexcl7days	Number of Inq Last 6 Months excl 7days. Excluding the last 7 days removes inquiries that are likely due to price comparision shopping.	Monotonically Increasing
NetFractionRevolvingBurden	Net Fraction Revolving Burden. This is revolving balance divided by credit limit	Monotonically Increasing
NetFractionInstallBurden	Net Fraction Installment Burden. This is installment balance divided by original loan amount	Monotonically Increasing
NumRevolvingTradesWBalance	Number Revolving Trades with Balance	No constraint
NumInstallTradesWBalance	Number Installment Trades with Balance	No constraint
NumBank2NatlTradesWHighUtilization	Number Bank/Natl Trades w high utilization ratio	Monotonically Increasing
PercentTradesWBalance	Percent Trades with Balance	No constraint
RiskPerformance	Paid as negotiated flag (12-36 Months). String of Good and Bad	Target

Questions that we ask

Data Scientists:

- What is the overall logic of the model in making decisions?
- Is the logic reasonable, so that we can deploy the model with confidence?

Loan Officers:

- Why is the model recommending this person's credit be approved or denied?
- How can I inform my decision to accept or reject a line of credit by looking at similar individuals?

Bank Customers:

- Why was my application rejected?
- What can I improve to increase the likelihood my application is accepted?

Field	Meaning	Monotonicity Constraint (with respect to probability of bad = 1)
ExternalRiskEstimate	Consolidated version of risk markers	Monotonically Decreasing
MSinceOldestTradeOpen	Months Since Oldest Trade Open	Monotonically Decreasing
MSinceMostRecentTradeOpen	Months Since Most Recent Trade Open	Monotonically Decreasing
AverageMInFile	Average Months in File	Monotonically Decreasing
NumSatisfactoryTrades	Number Satisfactory Trades	Monotonically Decreasing
NumTrades60Ever2DerogPubRec	Number Trades 60+ Ever	Monotonically Decreasing
NumTrades90Ever2DerogPubRec	Number Trades 90+ Ever	Monotonically Decreasing
PercentTradesNeverDelq	Percent Trades Never Delinquent	Monotonically Decreasing
MSinceMostRecentDelq	Months Since Most Recent Delinquency	Monotonically Decreasing
MaxDelq2PublicRecLast12M	Max Delq/Public Records Last 12 Months. See tab "MaxDelq" for each category	Values 0-7 are monotonically decreasing
MaxDelqEver	Max Delinquency Ever. See tab "MaxDelq" for each category	Values 2-8 are monotonically decreasing
NumTotalTrades	Number of Total Trades (total number of credit accounts)	No constraint
NumTradesOpeninLast12M	Number of Trades Open in Last 12 Months	Monotonically Increasing
PercentInstallTrades	Percent Installment Trades	No constraint
MSinceMostRecentInqexcl7days	Months Since Most Recent Inq excl 7days	Monotonically Decreasing
NumInqLast6M	Number of Inq Last 6 Months	Monotonically Increasing
NumInqLast6Mexcl7days	Number of Inq Last 6 Months excl 7days. Excluding the last 7 days removes inquiries that are likely due to price comparision shopping.	Monotonically Increasing
NetFractionRevolvingBurden	Net Fraction Revolving Burden. This is revolving balance divided by credit limit	Monotonically Increasing
NetFractionInstallBurden	Net Fraction Installment Burden. This is installment balance divided by original loan amount	Monotonically Increasing
NumRevolvingTradesWBalance	Number Revolving Trades with Balance	No constraint
NumInstallTradesWBalance	Number Installment Trades with Balance	No constraint
NumBank2NatlTradesWHighUtilization	Number Bank/Natl Trades w high utilization ratio	Monotonically Increasing
PercentTradesWBalance	Percent Trades with Balance	No constraint
RiskPerformance	Paid as negotiated flag (12-36 Months). String of Good and Bad	Target

Picking the Appropriate Fairness Metrics for One's User-persona

Data Scientist: Must ensure the model works appropriately before deployment

- Generalized Linear Rule Model (GLRM)

Loan Officer: Needs to assess the model's prediction and make the final judgement

- ProtoDash

Bank Customer: Wants to understand the reason of application result

- Contrastive Explanations Method

How ProtoDash helps the Loan Officer

Questions asked by Loan Officers:

- Why is the model recommending this person's credit be approved or denied?
- How can I inform my decision to accept or reject a line of credit by looking at similar individuals?

How ProtoDash works

- Works with an existing predictive model to show how the customer compares to others who have similar profiles and had similar repayment records to the model's prediction for the current customer, which helps to evaluate and predict the applicant's risk.

- Based on the model's prediction and the explanation for how it came to that recommendation, the Loan Officer can make a more informed decision.

Finding Similar Profiles in the Dataset Based on Outcome for a Customer

Alice
Approved

	Alice	Mia	Kate	Cala
Outcome	-	Paid	Paid	Paid
Similarity to Alice (from 0 to 1)	-	0.765	0.081	0.065
ExternalRiskEstimate	82	85	80	89
MSinceOldestTradeOpen	280	223	382	379
MSinceMostRecentTradeOpen	13	13	4	156
AverageMInFile	102	87	90	257
NumSatisfactoryTrades	22	23	21	3
NumTrades60Ever2DerogPubRec	0	0	0	0
NumTrades90Ever2DerogPubRec	0	0	0	0
PercentTradesNeverDelq	91	91	95	100
MSinceMostRecentDelq	26	26	69	0
MaxDelq2PublicRecLast12M	6	6	6	7
MaxDelqEver	6	6	6	8

Robert
Denied

	Robert	James	Danielle	Franklin
Outcome	-	Defaulted	Defaulted	Defaulted
Similarity to Robert (from 0 to 1)	-	0.690	0.114	0.108
ExternalRiskEstimate	78	71	72	69
MSinceOldestTradeOpen	82	95	166	193
MSinceMostRecentTradeOpen	5	1	12	12
AverageMInFile	54	43	74	167
NumSatisfactoryTrades	33	33	37	36
NumTrades60Ever2DerogPubRec	0	0	1	0
NumTrades90Ever2DerogPubRec	0	0	1	0
PercentTradesNeverDelq	100	100	95	100
MSinceMostRecentDelq	0	0	7	0
MaxDelq2PublicRecLast12M	7	7	4	7
MaxDelqEver	8	8	4	8

Note: Value is **highlighted** for similar profiles (columns) when it is same as that of given customer (second column)

Explanations for Data Scientists

Algorithm 1:
Boolean Rule Column Generation (BRCG):

simple OR-of-ANDs classification rules

Predict Y=0 if ANY of the following rules are satisfied, otherwise Y=1:

['ExternalRiskEstimate <= 75.00 AND NumSatisfactoryTrades <= 17.00', 'ExternalRiskEstimate <= 72.00 AND NumSatisfactoryTrades > 17.00']

Algorithm 2: Logistic Rule Regression (LRR):
weighted combinations of rules

Training accuracy: 0.7426718897744158

Test accuracy: 0.7260940032414911

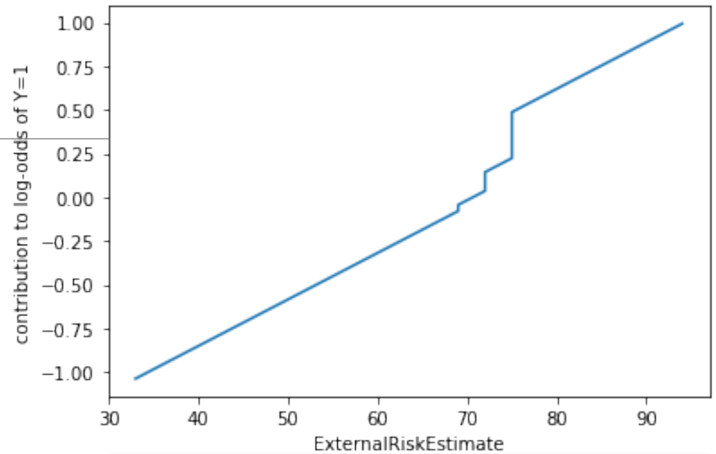
Probability of Y=1 is predicted as logistic(z) = 1 / (1 + exp(-z)) where z is a linear combination of the following rules/numerical features:

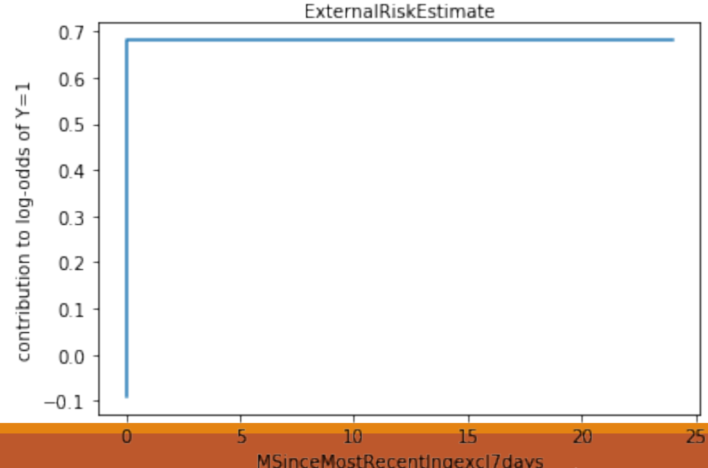
	rule/numerical feature	coefficient
0	(intercept)	-0.129693
1	MSinceMostRecentInqexcl7days > 0.00	0.680256
2	ExternalRiskEstimate	0.654176
3	NetFractionRevolvingBurden	-0.554147
4	NumSatisfactoryTrades	0.551635
5	NumInqLast6M	-0.463194
6	NumBank2NatlTradesWHighUtilization	-0.448368
7	AverageMInFile <= 52.00	-0.43437
8	NumRevolvingTradesWBalance <= 5.00	0.421518

Explanations for Data Scientists

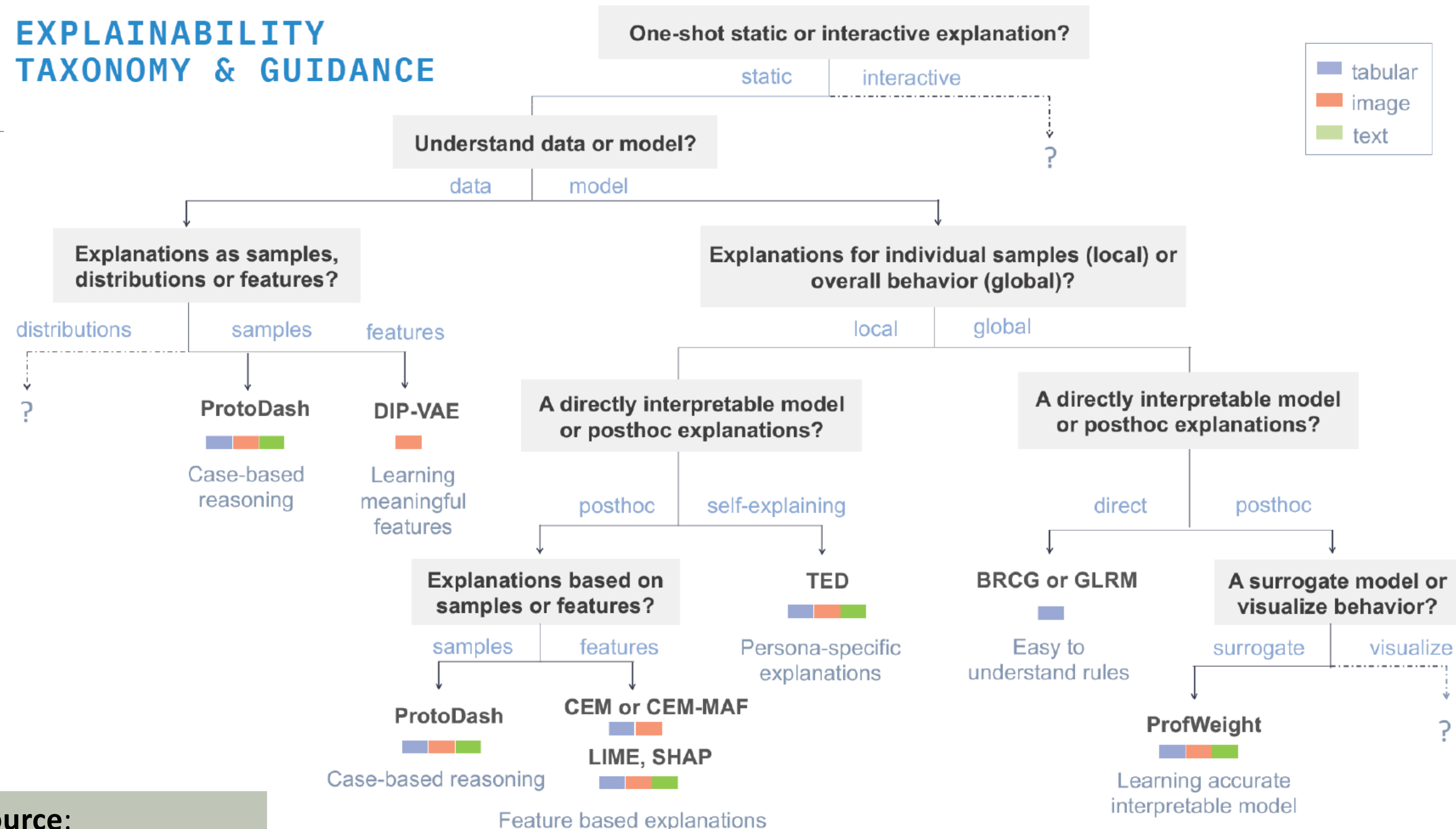
Algorithm 2: Logistic Rule Regression (LRR):
weighted combinations of rules

	rule/numerical feature	coefficient
0	(intercept)	-0.129693
1	MSinceMostRecentInqexcl7days > 0.00	0.680256
2	ExternalRiskEstimate	0.654176
3	NetFractionRevolvingBurden	-0.554147
4	NumSatisfactoryTrades	0.551635
5	NumInqLast6M	-0.463194
6	NumBank2NatlTradesWHighUtilization	-0.448368
7	AverageMInFile <= 52.00	-0.43437
8	NumRevolvingTradesWBalance <= 5.00	0.421518





A Spectrum of Explanations in AIX360



Slide Source:
Vera Liao's XAI talk 2020

Emerging Support for Explanation in AI Offerings

Toolkit	Data Explanation	Directly interpretable	Global post-hoc	Local/inspection post-hoc	Customizable explanation	Metrics
AIX 360	ProtoDash, DIP-VAE	BRCG, GLRM	ProfWeight	LIME, SHAP, CEM, CEM-MAF, ProtoDash	TED	Faithfulness, Monotonicity
Seldon Alibi			✓	✓		
Oracle Skater		✓	✓	✓		
H2o		✓	✓	✓		
Microsoft Interpret		✓	✓	✓		
DALEX			✓	✓		
Ethical ML			✓			

Slide Source:
Vera Liao's XAI talk 2020

Automate AI (AutoAI)

- Also called AutoML, Automated Data Science
- Objectives
  - Automate the mundane tasks in ML pipeline
  - Improve effectiveness (e.g., accuracy)

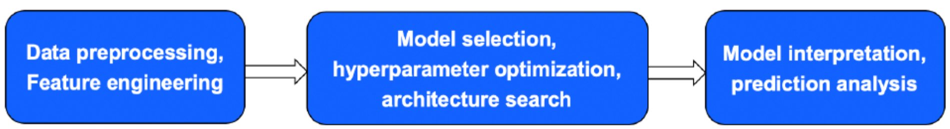


Image source: Towards Automated Machine Learning: Evaluation and Comparison of AutoML Approaches and Tools

AutoAI – Task Details

- Data preprocessing
  - Data cleaning
  - Missing data imputation
  - Data transformation (e.g., categorical, time) and normalization
- Feature engineering/ selection
  - Drop dependent features
  - Create new features

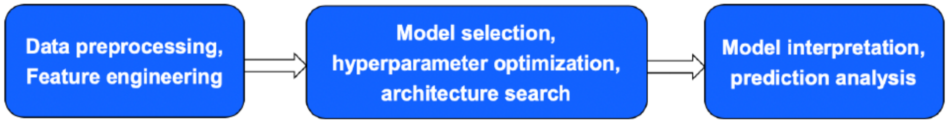


Image source: Towards Automated Machine Learning: Evaluation and Comparison of AutoML Approaches and Tools

AutoAI – Task Details

- Model selection
  - Single
  - Ensemble
- Hyper-parameter via search strategies
- Architecture search
  - Neural network (layers), weight sharing
  - ML pipelines

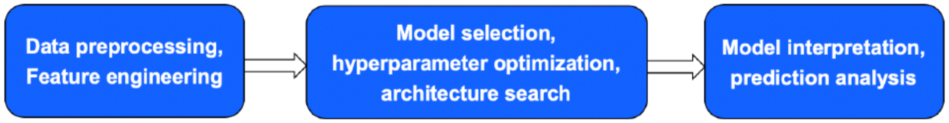


Image source: Towards Automated Machine Learning: Evaluation and Comparison of AutoML Approaches and Tools

AutoAI – Why Do It and Why Don't

- Pros
  - Improves performance (accuracy)
  - Speeds up insight generation process
- Cons
  - Over-reliance, especially in dynamic environment, makes on miss issues
  - High resource consumption (memory, energy)

Code Examples

- AutoAI
  - https://github.com/biplav-s/course-d2d-ai/tree/main/sample-code/l12-explanability-autoai

- We will see:
  - Data profiling
  - Semi-automation – Lale
  - Automated – auto-sklearn

Auto AI References

1. Feature Engineering for Predictive Modeling using Reinforcement Learning, Udayan Khurana, Horst Samulowitz, Deepak Turaga, AAAI 2018 paper, pre-print https://arxiv.org/pdf/1709.07150.pdf
2. 2. Learning Feature Engineering for Classification, Fatemeh Nargesian, Horst Samulowitz, Udayan Khurana, Elias B. Khalil, Deepak Turaga, IJCAI 2017, https://www.ijcai.org/proceedings/2017/0352.pdf
3. 3. Cracking open the black box of automated machine learning, http://news.mit.edu/2019/atmseer-machine-learning-black-box-0531
4. Though AI Outperforms Humans in Building AI, Human-AI Collaboration, The Future Of Data Science, Dakuo Wang et al, 2020, https://arxiv.org/abs/1909.02309

Lecture 12: Concluding Comments

- Generating explanations
  - LIME
  - AIX 360
  - Which methods work under what conditions?

- AutoAI
  - For removing mundane steps
  - Improving model performance

Concluding Segment

About Next Lecture – Lecture 13

Auto AI Paper

- Towards Automated Machine Learning: Evaluation and Comparison of AutoML Approaches and Tools
  - https://arxiv.org/abs/1908.05557, 2019

Reading Group Allocation

a	Brendan	Curran
b	Sarayu	Das
c	Marilyn	Gartley
d	Vedant	Khandelwal
e	Vishal	Pallagani
a	Avineet Kumar	Singh
b	Ahad Hasan	Tanim
c	James	Thompson
d	Rohit	Naini
e	Terric	Taylor

Quiz 2

- Classification
- Clustering
- Bonus question

Lecture 13: Time Series Analysis

- AutoAI paper discussion
- Time series - methods