SYNTHETIC DATA GENERATION WORKSHOP

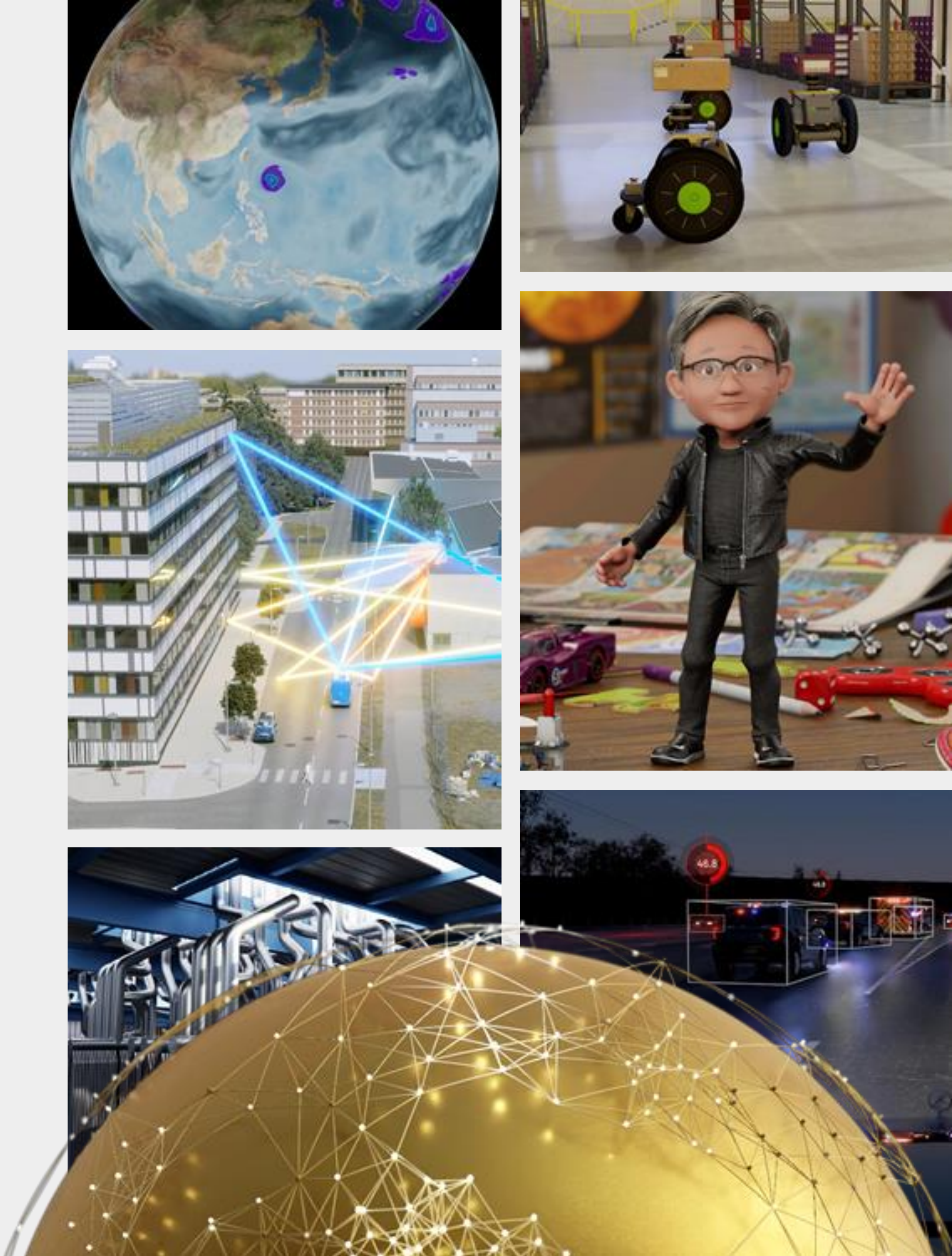
Nvidia Deep Learning Institute

NVIDIA GTC

BLOOMBERG
GENERATING AND ASSESSING SYNTHETIC TRANSACTION DATA

COHEN & STEERS
TRANSFORMER MODELS FOR TIMESERIES FORECASTING OF INFLATION AND MARKET REGIMES

https://www.nvidia.com/en-us/on-demand



MARCH 21—24, 2022 | www.nvidia.com/GTC

OUTLINE

- What is synthetic data generation and what motivates its use?
- Alternative methods
- Synthetic Data Generation Lab
  - ETL/Preprocessing
  - Model Training
  - Inferencing
  - Evaluation
- Next steps and Applications

SYNTHETIC DATA GENERATION

SYNTHETIC DATA GENERATION

What is synthetic data generation?

Any technique to augment existing datasets or create new ones

Why is this important?

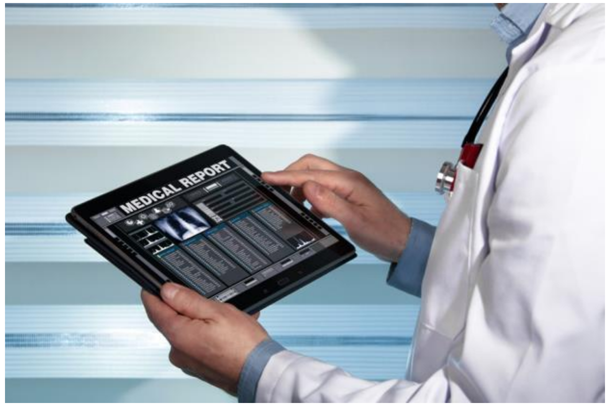
- Necessary for increasing model robustness and accuracy
- Maintain privacy
- Increase data diversity
- Sharing data with external stakeholders

The focus for this talk is on <u>tabular</u> synthetic data generation

APPLICATIONS

Fraud Detection and Cybersecurity

Backtesting equities



Electronic Medical Records

FEATURES OF SYNTHETIC DATA

CREDIT CARD DATA

Real – 24M rows

user	card	amount	date	year	month	day	hour	minute	use chip	merchant name	merchant city	merchant state	zip	mcc	errors	is fraud
791	1	68.00	2018-01-02 09:10:00	2018	1	2	9	10	Swipe Transaction	12345536	New York	NY	10017	8005	<NA>	0
1572	0	572.42	2018-04-12 07:11:00	2018	4	12	7	11	Chip Transaction	49908535	Princeton	NJ	19406	5634	<NA>	0
2718	7	123.10	2019-01-04 10:14:00	2019	1	4	10	14	Chip Transaction	43211536	Beverly Hills	CA	90210	4800	<NA>	0
21	2	42.04	2020-06-23 11:18:00	2020	6	23	11	18	Swipe Transaction	65423006	Burke	VA	22015	5604	<NA>	0
1001	1	5000.00	2020-11-03 01:22:00	2020	11	3	1	22	Online Transaction	75434546	<NA>	<NA>	<NA>	1234	<NA>	1

Synthetic – 42M rows

user	card	amount	date	year	month	day	hour	minute	use chip	merchant name	merchant city	merchant state	zip	mcc	errors	is_fraud
1010	3	68.64	2019-07-22 12:43:00	2019	7	22	12	43	Chip Transaction	2027553650310142703	Boxford	MA	01921	5541	<NA>	0
142	0	2.21	2004-10-07 06:08:00	2004	10	7	6	8	Swipe Transaction	-6571010470072147219	Seattle	WA	98102	5499	<NA>	0
1037	1	24.32	2014-11-23 17:41:00	2014	11	23	17	41	Swipe Transaction	3959361429988996167	Tucson	AZ	85719	5912	<NA>	0
1734	0	29.60	2004-11-26 22:20:00	2004	11	26	22	20	Swipe Transaction	-4530600671233798827	Menlo Park	CA	94025	5812	<NA>	0
118	1	60.72	2018-11-16 21:53:00	2018	11	16	21	53	Chip Transaction	4751695835751691036	Anaheim	CA	92801	5814	<NA>	0

Tabular Transformers for Modeling Multivariate Time Series (arxiv.org)

IMPORTANT FEATURES OF SYNTHETIC DATA

City, State
✗ San Francisco, NY
✓ San Francisco, CA

- **Representative** of underlying real data
  - Real and Synthetic Data have same columns
  - Data are drawn from a similar distribution
  - Synthetic data accurately represents global trends, and local trends in the real data
  - Relevant cross-column categorical features (i.e. city, state)
- **Privacy-focused**
  - Does not leak information about specific entities in the real data
- **Conditionally Generated**
  - Generate new data based on a provided “context”
  - Generate new edge case data

SAMPLE OF CURRENT APPROACHES

Classical:

- Oversampling - ex. SMOTE (Synthetic minority oversampling)
- Bagging - Bootstrap aggregation
- PCA - Principal component analysis

Deep Learning:

- Variational Autoencoders (VAEs)
- Generative Adversarial Networks (GANs)
- Transformers (current focus)

Practical Issues:

- **Loss** of time-based trends
- **Add bias** to a model by reusing existing data
- **Loss** of privacy
- **Interpolation may not make sense** for certain data (ex. categorical data such as zip codes)
- **Capturing associations** across columns can be hard (ex. zip codes associated with city/state)
- **Catastrophic Forgetting** - the model forgets previous information upon learning new information
- **Posterior collapse** - the model only outputs a single value.

WHAT DOES MODE COLLAPSE LOOK LIKE?

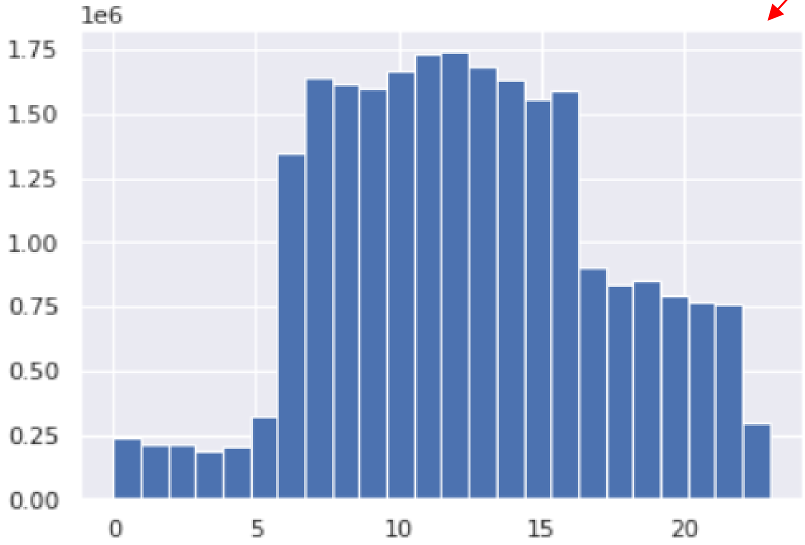
VAE MODE COLLAPSE

Generated Data

	card	day	errors_Bad CVV	errors_Bad Card Number	errors_Bad Expiration	errors_Bad PIN	errors_Bad Zipcode	errors_Insufficient Balance	errors_Technical Glitch	hour	mcc	merchant city	merchant name	merchant state	merchant_city_state_zip	minute	month	num_cards_per_user	use chip	year	zip	amoun
0	1	25	0	0	0	0	0	0	0	11	891	1825	39991	151	13508	38	7	4	0	21	22238	43.48357
1	0	7	0	0	0	0	0	0	0	12	609	9622	22204	0	7002	53	7	3	0	27	2119	42.92538
2	0	25	0	0	0	0	0	0	0	12	631	6244	5248	151	7002	36	8	3	0	25	0	43.33191
3	1	21	0	0	0	0	0	0	0	11	612	786	7777	192	13508	24	1	4	0	24	9456	44.39045
4	0	7	0	0	0	0	0	0	0	12	627	5054	3831	0	7002	31	8	3	0	23	3604	43.10829

```
traindf.hour.to_pandas().hist(bins=24)
```

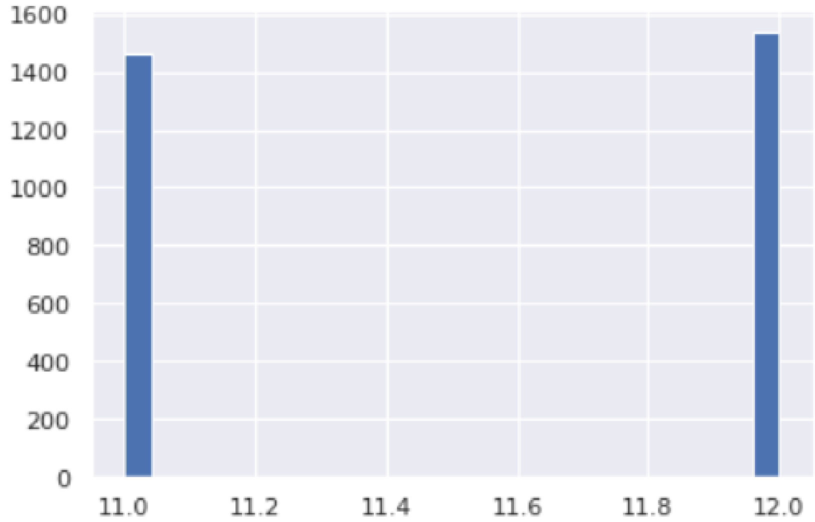
```
<AxesSubplot:>
```



Real Distribution

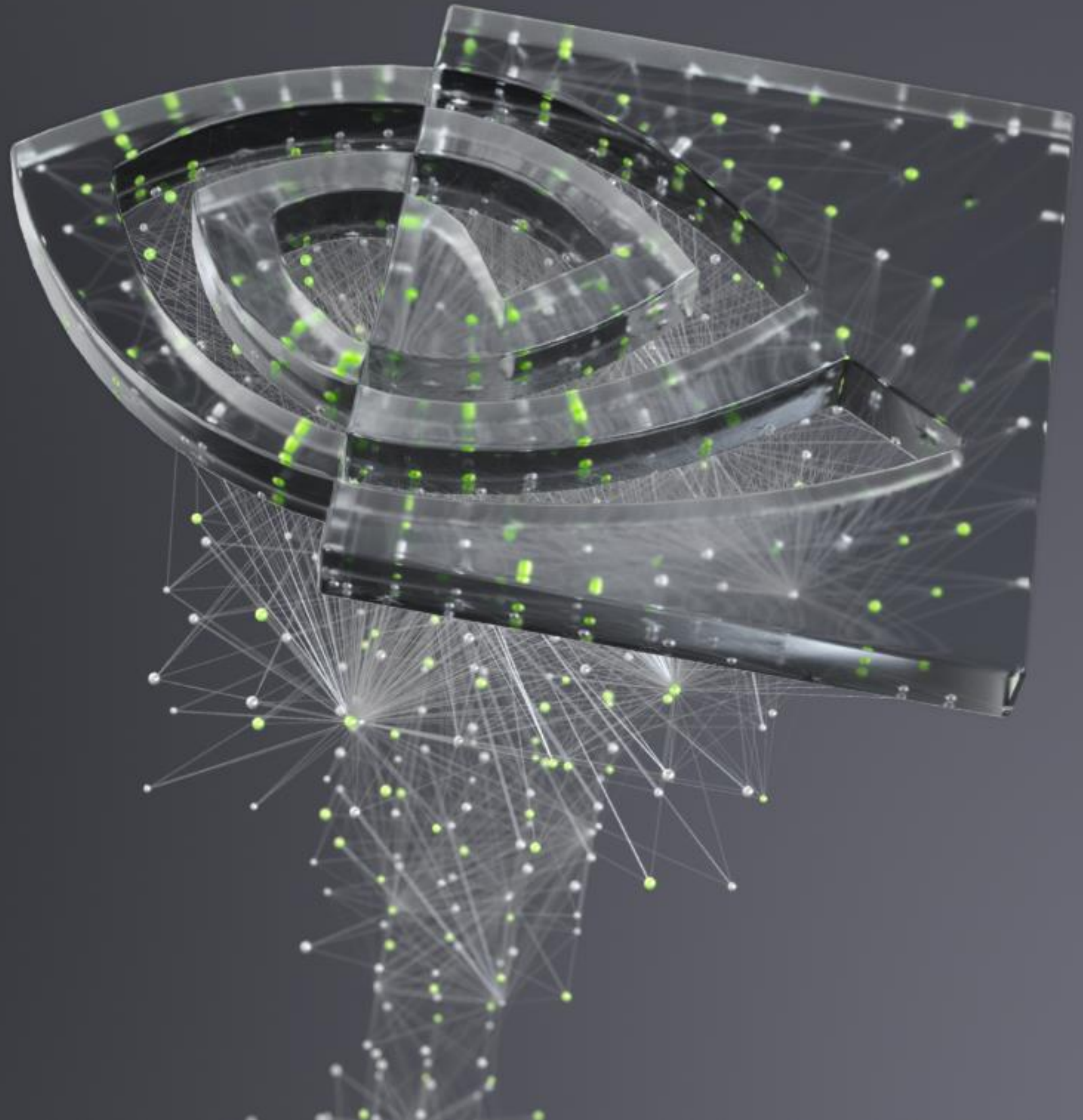
```
output.hour.to_pandas().hist(bins=24)
```

```
<AxesSubplot:>
```



Generated Distribution

SYNTHETIC DATA WORKFLOW



SYNTHETIC DATA WORKFLOW

Overview to generate synthetic data

ETL/Preprocess → Model Training → Inference → Evaluation

For our lab we will use a credit card payments dataset to demonstrate this process

ETL / PREPROCESSING

SCALE OUT PYTHON TOOLS WITH RAPIDS + DASK

DISTRIBUTE & ACCELERATE COMPUTATION FOR PRODUCTION WORKLOADS

Scale Up / Accelerate

RAPIDS

Accelerates PyData on NVIDIA GPUs

NumPy -> CuPy/PyTorch/..
Pandas -> cuDF
Scikit-Learn -> cuML
Numba -> Numba

RAPIDS + DASK

Distributes and accelerates PyData

Can be distributed across Multi-GPU on single node (DGX) or across a cluster

Provides easy to use tooling enabling HPC-level performance

PYDATA

Provides accessible, easy to use tooling

NumPy, Pandas, Scikit-Learn, Numba and many more

Single CPU core, in-memory data

DASK

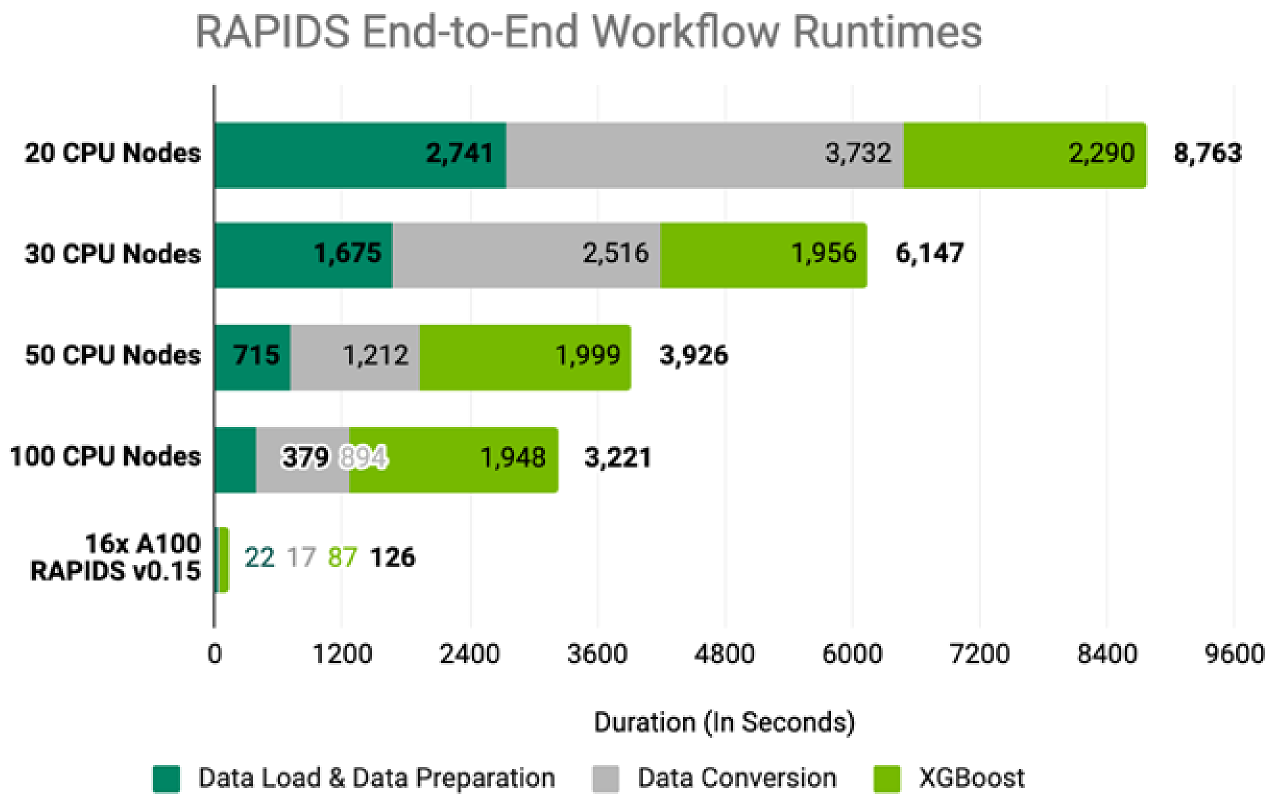
Distributes PyData across multiple cores

NumPy -> Dask Array
Pandas -> Dask DataFrame
Scikit-Learn -> Dask-ML
… -> Dask Futures

Scale Out / Parallelize

LIGHTNING-FAST END-TO-END PERFORMANCE

REDUCING DATA SCIENCE PROCESSES FROM HOURS TO SECONDS



16

A100s Provide More Power than 100 CPU Nodes

70x

Faster Performance than Similar CPU Configuration

20x

More Cost-Effective than Similar CPU Configuration

*CPU approximate to n1-highmem-8 (8 vCPUs, 52GB memory) on Google Cloud Platform. TCO calculations-based on Cloud instance costs.

CHALLENGES OF DIRECTLY APPLYING NLP TOKENIZER TO TABULAR DATA

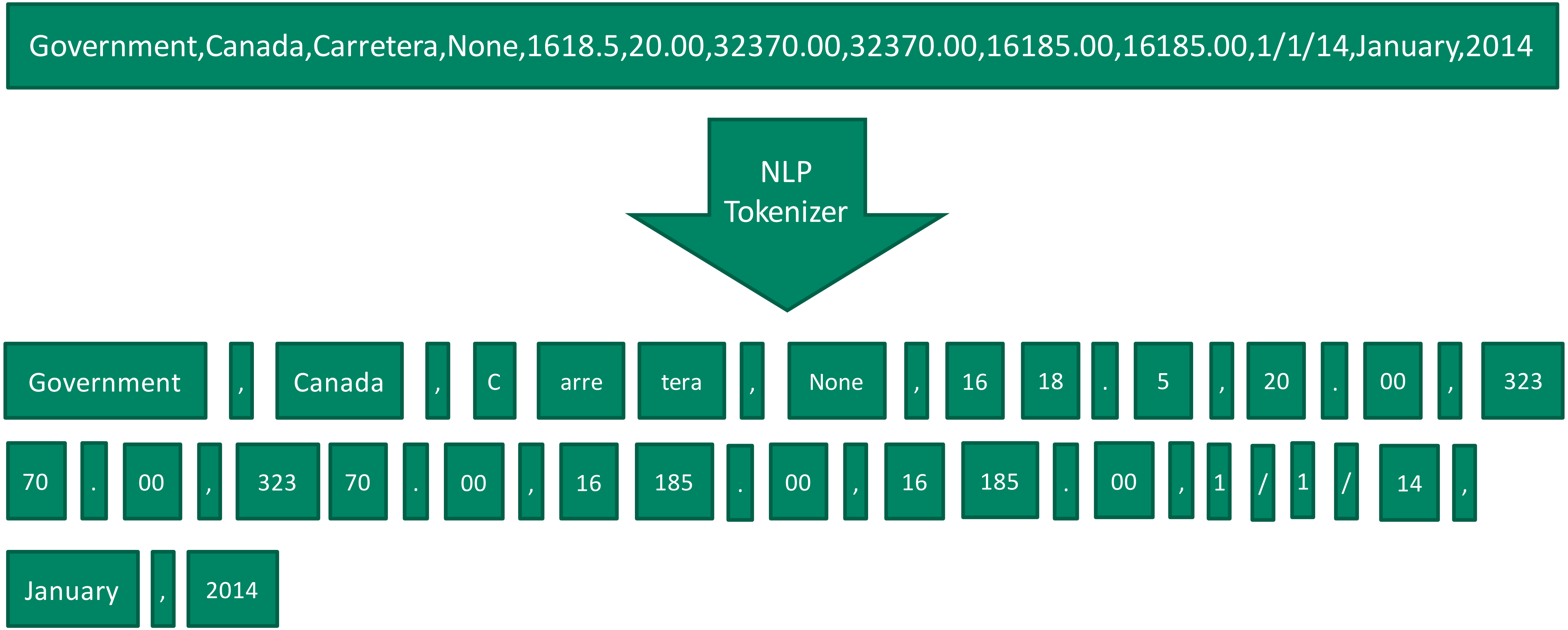
TABULAR DATA IS STRUCTURED

Segment	Country	Product	Discount Band	Units Sold	Sale Price	Gross Sales	Sales	COGS	Profit	Date	Month Name	Year
Government	Canada	Carretera	None	1618.5	$ 20.00	$ 32,370.00	$ 32,370.00	$ 16,185.00	$ 16,185.00	1/1/14	January	2014
Government	Germany	Carretera	None	1321	$ 20.00	$ 26,420.00	$ 26,420.00	$ 13,210.00	$ 13,210.00	1/1/14	January	2014
Midmarket	France	Carretera	None	2178	$ 15.00	$ 32,670.00	$ 32,670.00	$ 21,780.00	$ 10,890.00	6/1/14	June	2014
Midmarket	Germany	Carretera	None	888	$ 15.00	$ 13,320.00	$ 13,320.00	$ 8,880.00	$ 4,440.00	6/1/14	June	2014
Midmarket	Mexico	Carretera	None	2470	$ 15.00	$ 37,050.00	$ 37,050.00	$ 24,700.00	$ 12,350.00	6/1/14	June	2014
Government	Germany	Carretera	None	1513	$ 350.00	$ 529,550.00	$ 529,550.00	$ 393,380.00	$ 136,170.00	12/1/14	December	2014
Midmarket	Germany	Montana	None	921	$ 15.00	$ 13,815.00	$ 13,815.00	$ 9,210.00	$ 4,605.00	3/1/14	March	2014
Channel Partners	Canada	Montana	None	2518	$ 12.00	$ 30,216.00	$ 30,216.00	$ 7,554.00	$ 22,662.00	6/1/14	June	2014
Government	France	Montana	None	1899	$ 20.00	$ 37,980.00	$ 37,980.00	$ 18,990.00	$ 18,990.00	6/1/14	June	2014
Channel Partners	Germany	Montana	None	1545	$ 12.00	$ 18,540.00	$ 18,540.00	$ 4,635.00	$ 13,905.00	6/1/14	June	2014
Midmarket	Mexico	Montana	None	2470	$ 15.00	$ 37,050.00	$ 37,050.00	$ 24,700.00	$ 12,350.00	6/1/14	June	2014
Enterprise	Canada	Montana	None	2665.5	$ 125.00	$ 333,187.50	$ 333,187.50	$ 319,860.00	$ 13,327.50	7/1/14	July	2014
Small Business	Mexico	Montana	None	958	$ 300.00	$ 287,400.00	$ 287,400.00	$ 239,500.00	$ 47,900.00	8/1/14	August	2014
Government	Germany	Montana	None	2146	$ 7.00	$ 15,022.00	$ 15,022.00	$ 10,730.00	$ 4,292.00	9/1/14	September	2014
Enterprise	Canada	Montana	None	345	$ 125.00	$ 43,125.00	$ 43,125.00	$ 41,400.00	$ 1,725.00	10/1/13	October	2013
Midmarket	United States of America	Montana	None	615	$ 15.00	$ 9,225.00	$ 9,225.00	$ 6,150.00	$ 3,075.00	12/1/14	December	2014
Government	Canada	Paseo	None	292	$ 20.00	$ 5,840.00	$ 5,840.00	$ 2,920.00	$ 2,920.00	2/1/14	February	2014
Midmarket	Mexico	Paseo	None	974	$ 15.00	$ 14,610.00	$ 14,610.00	$ 9,740.00	$ 4,870.00	2/1/14	February	2014
Channel Partners	Canada	Paseo	None	2518	$ 12.00	$ 30,216.00	$ 30,216.00	$ 7,554.00	$ 22,662.00	6/1/14	June	2014
Government	Germany	Paseo	None	1006	$ 350.00	$ 352,100.00	$ 352,100.00	$ 261,560.00	$ 90,540.00	6/1/14	June	2014
Channel Partners	Germany	Paseo	None	367	$ 12.00	$ 4,404.00	$ 4,404.00	$ 1,101.00	$ 3,303.00	7/1/14	July	2014
Government	Mexico	Paseo	None	883	$ 7.00	$ 6,181.00	$ 6,181.00	$ 4,415.00	$ 1,766.00	8/1/14	August	2014
Midmarket	France	Paseo	None	549	$ 15.00	$ 8,235.00	$ 8,235.00	$ 5,490.00	$ 2,745.00	9/1/13	September	2013
Small Business	Mexico	Paseo	None	788	$ 300.00	$ 236,400.00	$ 236,400.00	$ 197,000.00	$ 39,400.00	9/1/13	September	2013
Midmarket	Mexico	Paseo	None	2472	$ 15.00	$ 37,080.00	$ 37,080.00	$ 24,720.00	$ 12,360.00	9/1/14	September	2014
Government	United States of America	Paseo	None	1143	$ 7.00	$ 8,001.00	$ 8,001.00	$ 5,715.00	$ 2,286.00	10/1/14	October	2014
Government	Canada	Paseo	None	1725	$ 350.00	$ 603,750.00	$ 603,750.00	$ 448,500.00	$ 155,250.00	11/1/13	November	2013
Channel Partners	United States of America	Paseo	None	912	$ 12.00	$ 10,944.00	$ 10,944.00	$ 2,736.00	$ 8,208.00	11/1/13	November	2013
Midmarket	Canada	Paseo	None	2152	$ 15.00	$ 32,280.00	$ 32,280.00	$ 21,520.00	$ 10,760.00	12/1/13	December	2013

NVIDIA.

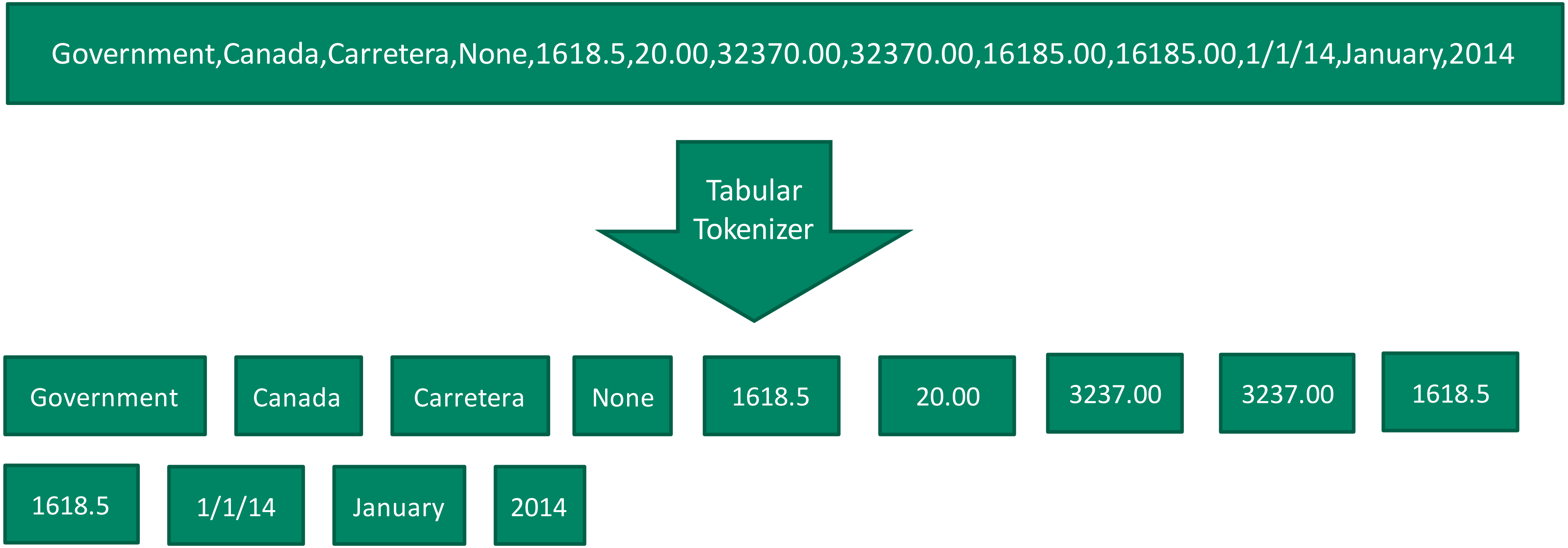
CHALLENGES OF DIRECTLY APPLYING NLP TOKENIZER TO TABULAR DATA

NLP tokenizer has no table structure information



SOLUTION: SPECIAL TABULAR TOKENIZER

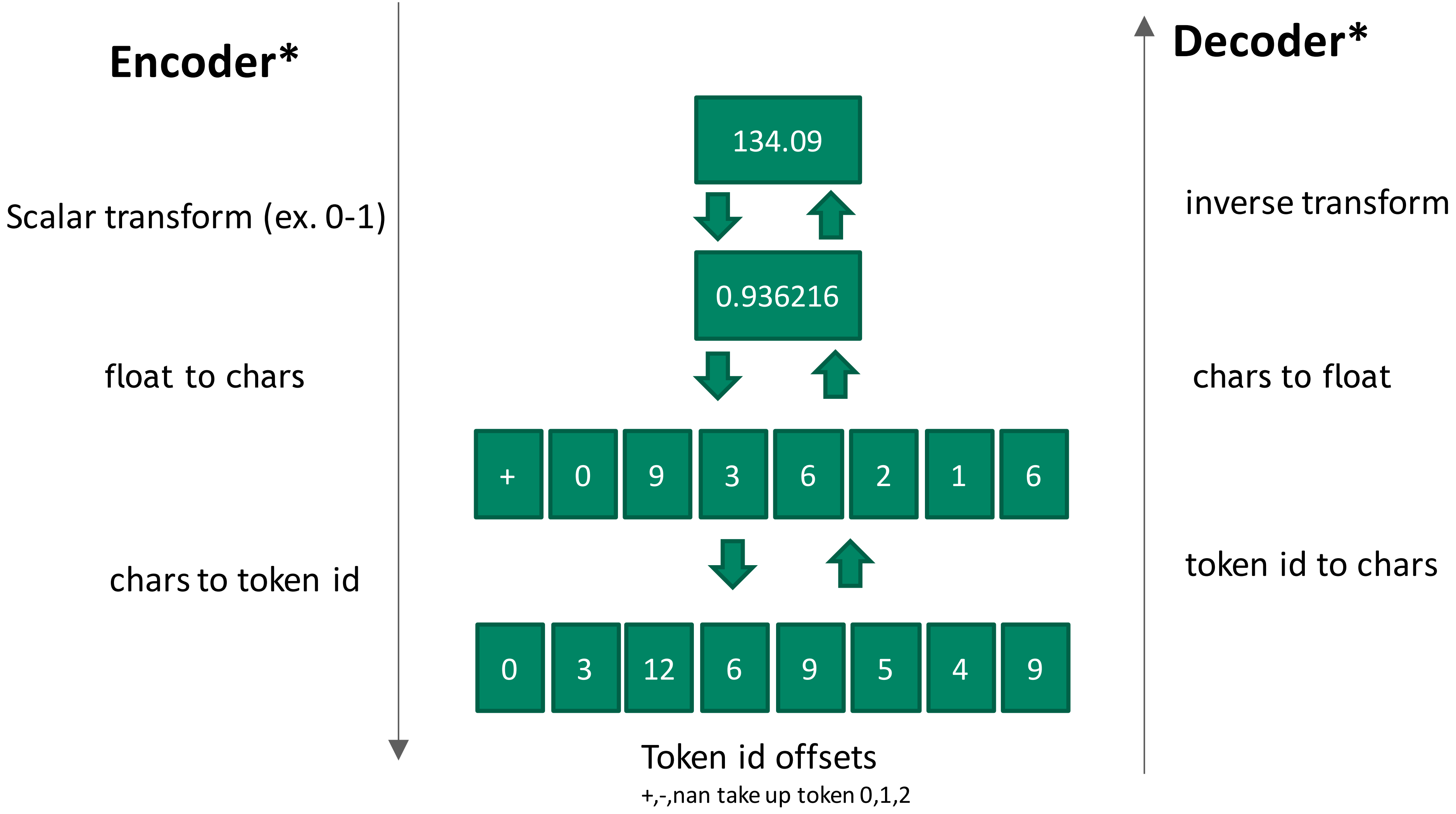
Tokenizer accounts for the table’s structural information



[Tabular transformer for modeling multivariate time series. Padhi et al 2021]()

SPECIAL FLOAT NUMBER ENCODER/DECODER

Encodes numeric value to tokens, and decodes back to numeric value



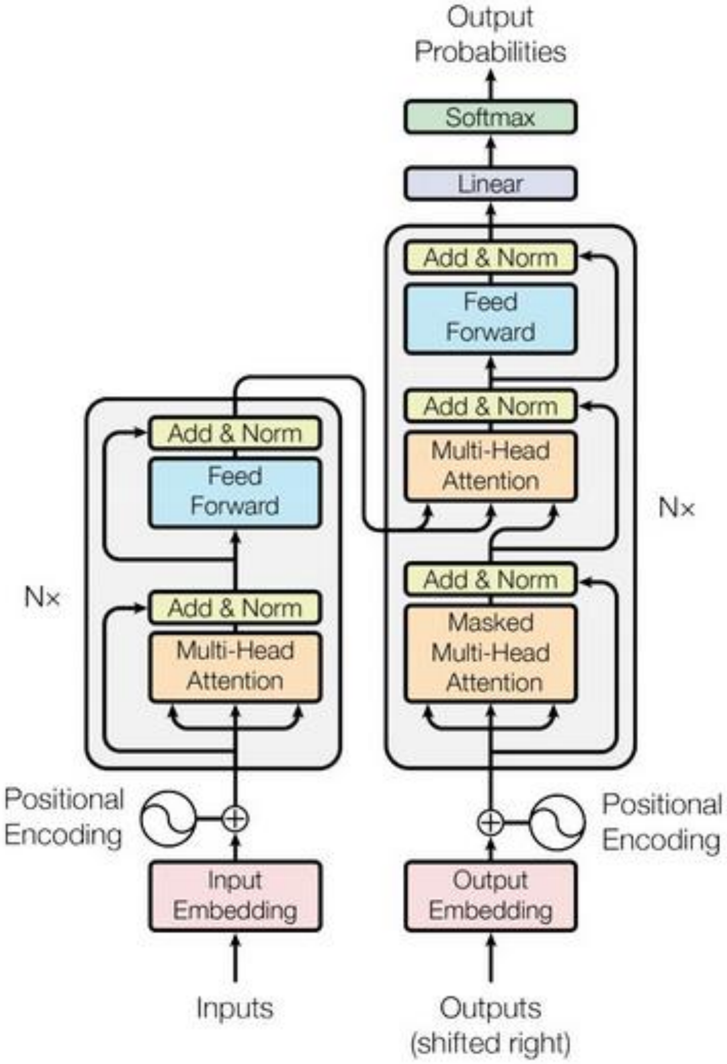
* column meta data (digits/min/max/transform) precalculated

Some Terms we will use in the rest of the workshop

- **Transformers** – Deep learning model architecture that is at the core of the state of the art in NLP tasks. Has and encoder / decoder component.

- **GPT** – the decoder component of the transformer model. First made popular by OpenAI GPT model that had amazing results in generative NLP tasks.

- **Megatron** – Nvidia's framework for accelerating multi-billion parameter transformer networks



ACTION
PLEASE RUN NOTEBOOKS

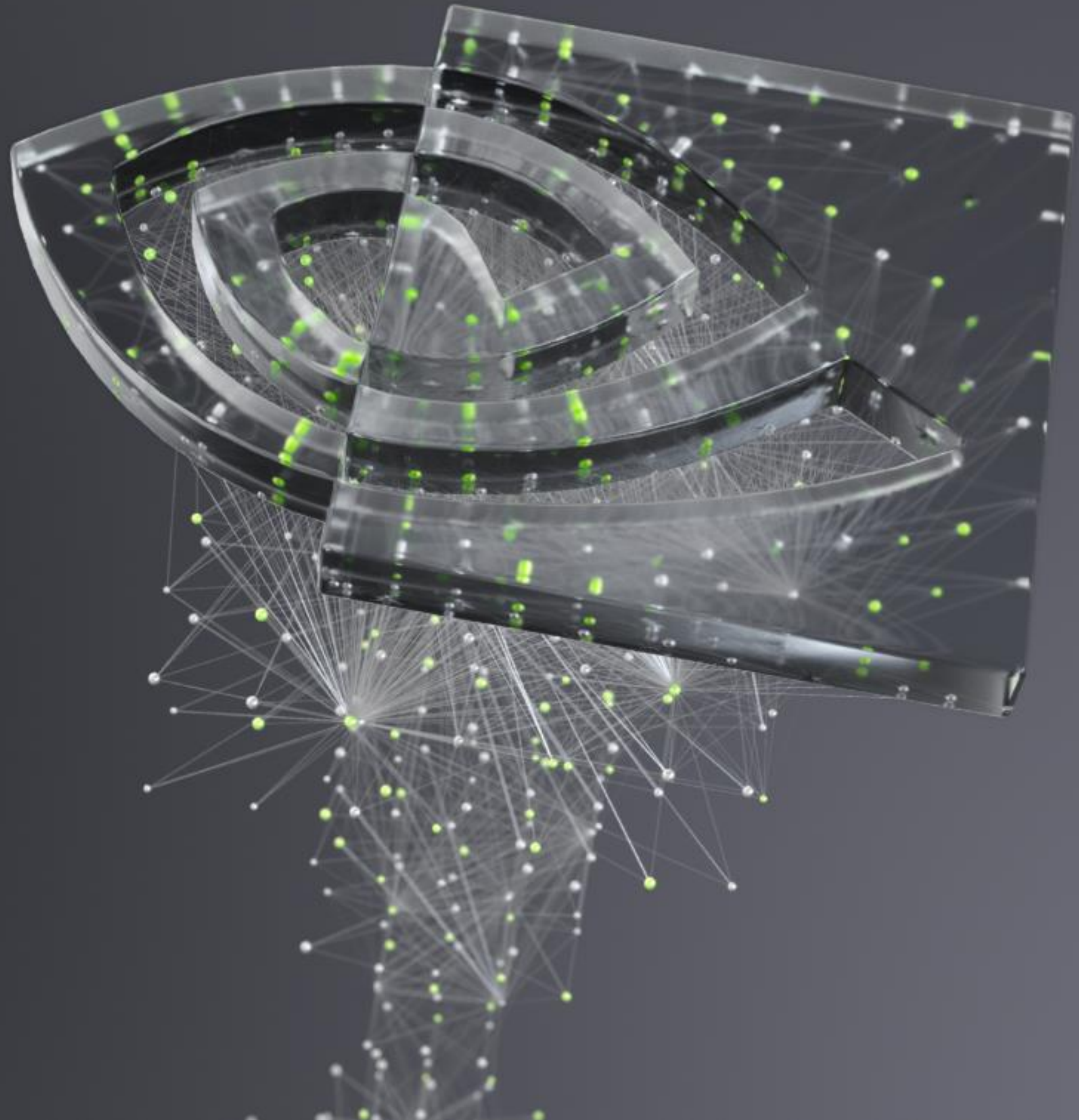
ETL/Preprocess → Model Training → Inference → Evaluation

0_Primer.ipynb
1_Megatron_Preprocessing.ipynb

Estimated time: 35 min

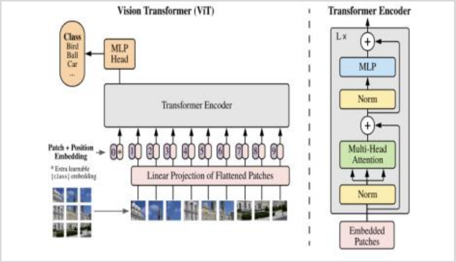
MODEL TRAINING

Transformers with NVIDIA Megatron

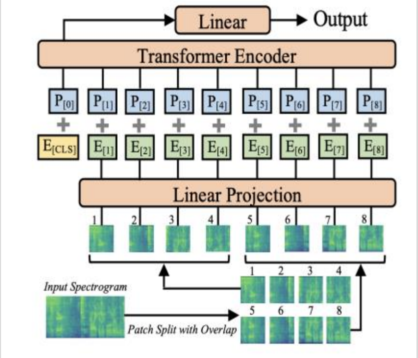


TRANSFORMER USED FOR MULTIPLE DOMAINS

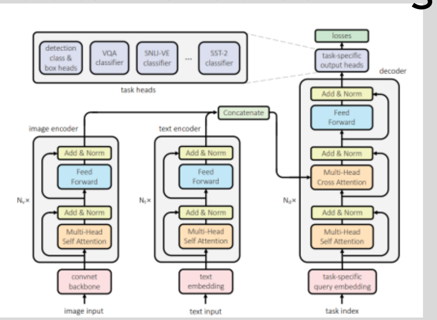
Image



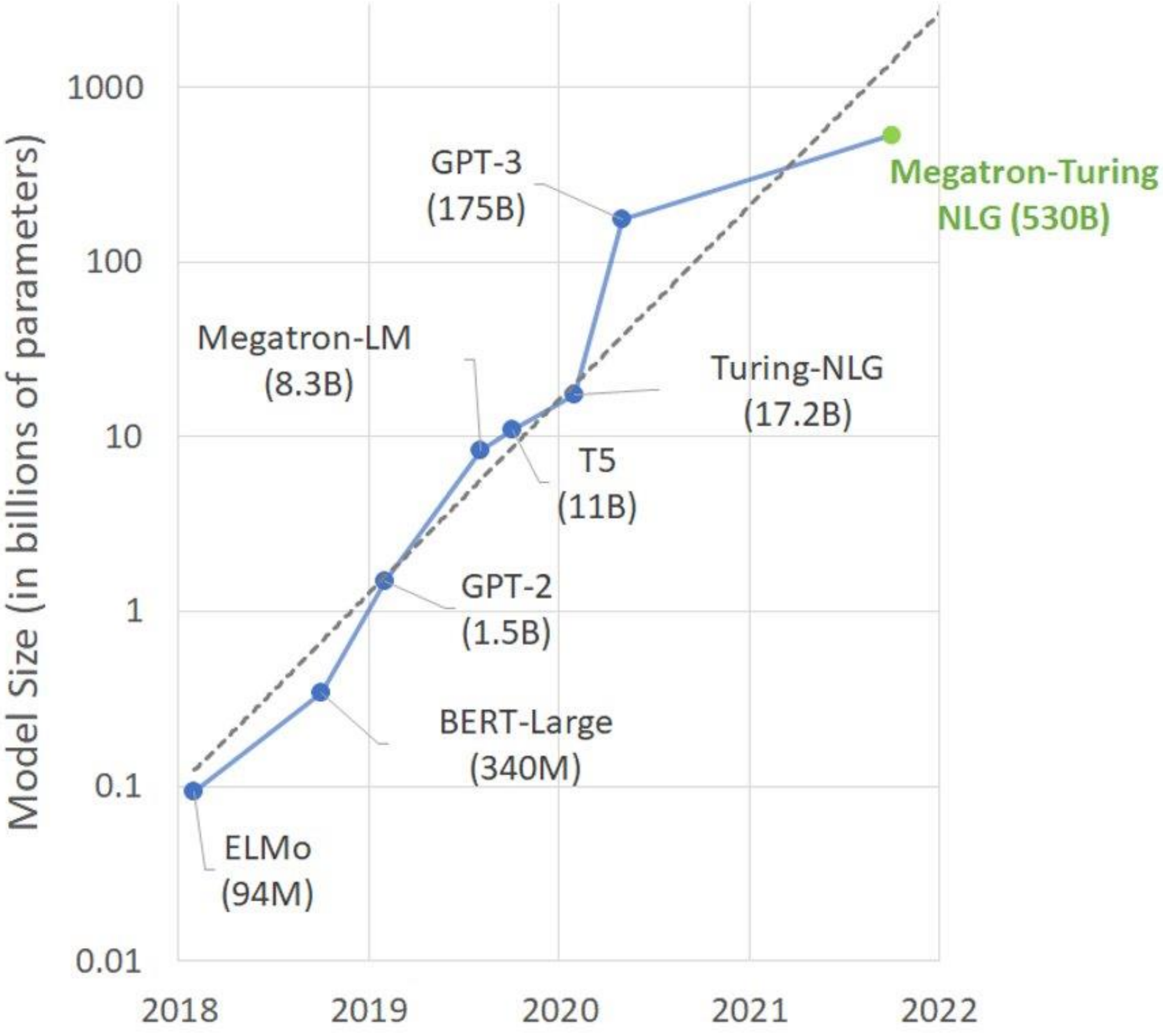
Audio

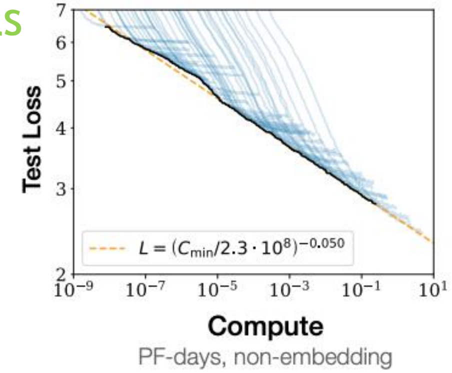


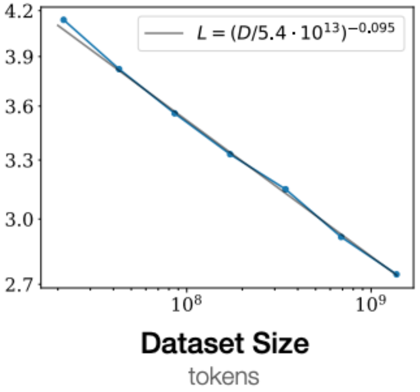
Multimodal learning

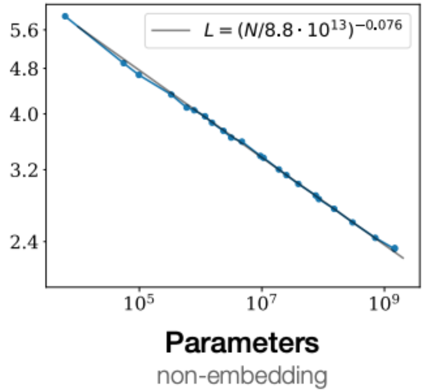


Explosive Growth of Model Sizes Yields More Accurate Models





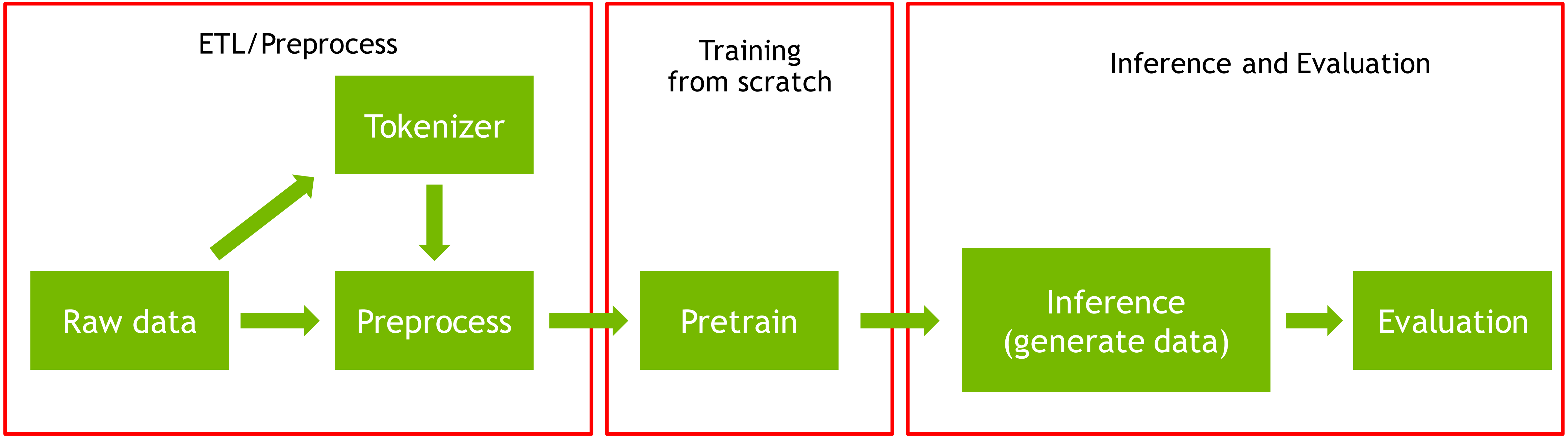




[2001.08361] Scaling Laws for Neural Language Models (arxiv.org)

MEGATRON GPT MODEL TRAINING PIPELINE

Using the Megatron Framework to train an NLP Model



ACTION
PLEASE RUN NOTEBOOKS

ETL/Preprocess → Model Training → Inference → Evaluation

Stop

0_Primer.ipynb
1_Megatron_Preprocessing

Start

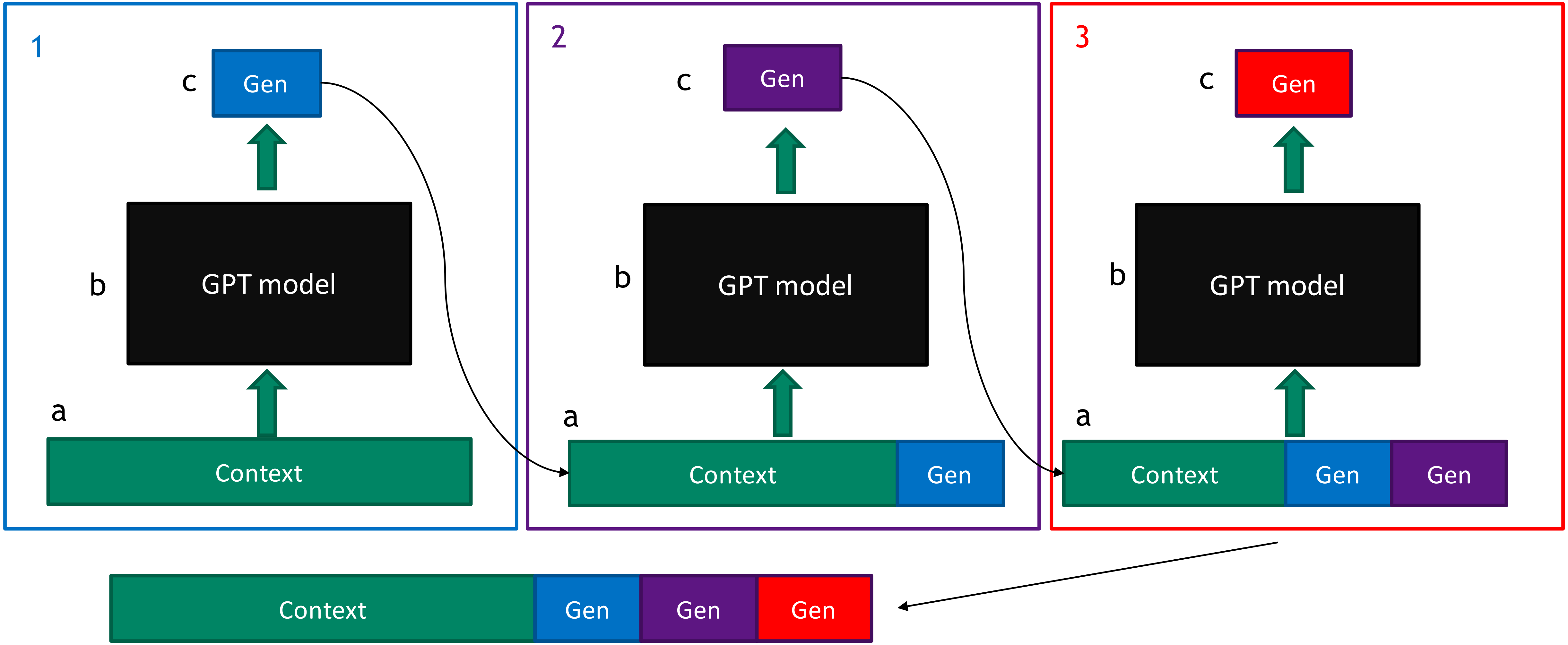
2_Training_Megatron.ipynb
2b_Tensorboard.ipynb

Estimated time: 20 min

INFERENCE AND EVALUATION

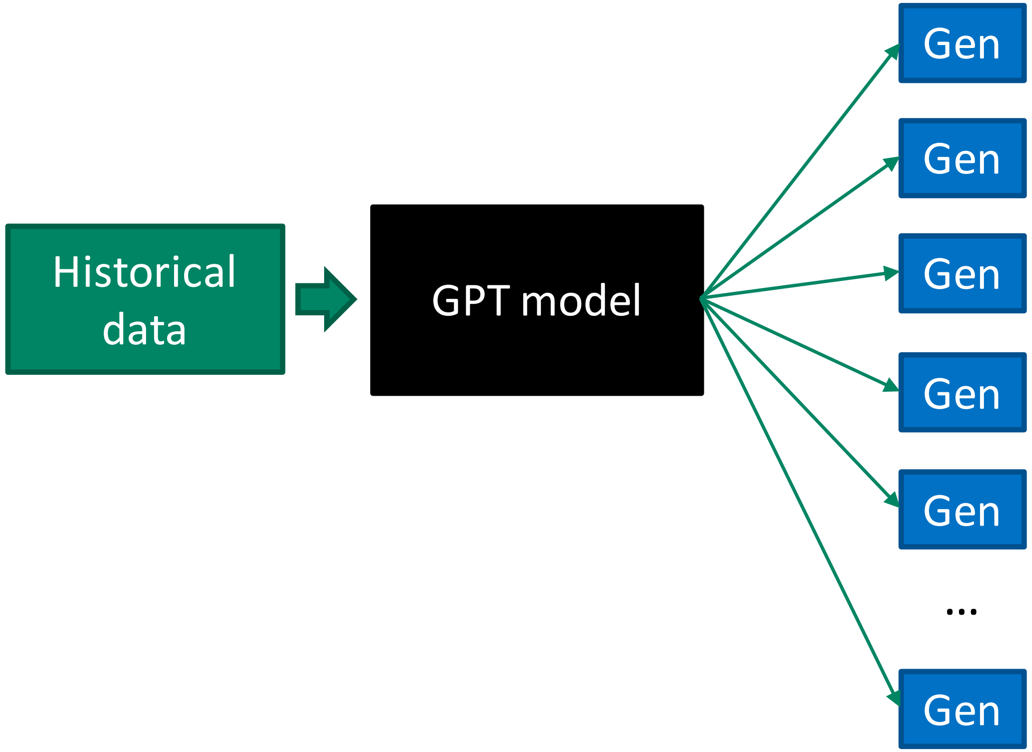
CONDITIONAL DATA GENERATION FOR LONG SEQUENCES

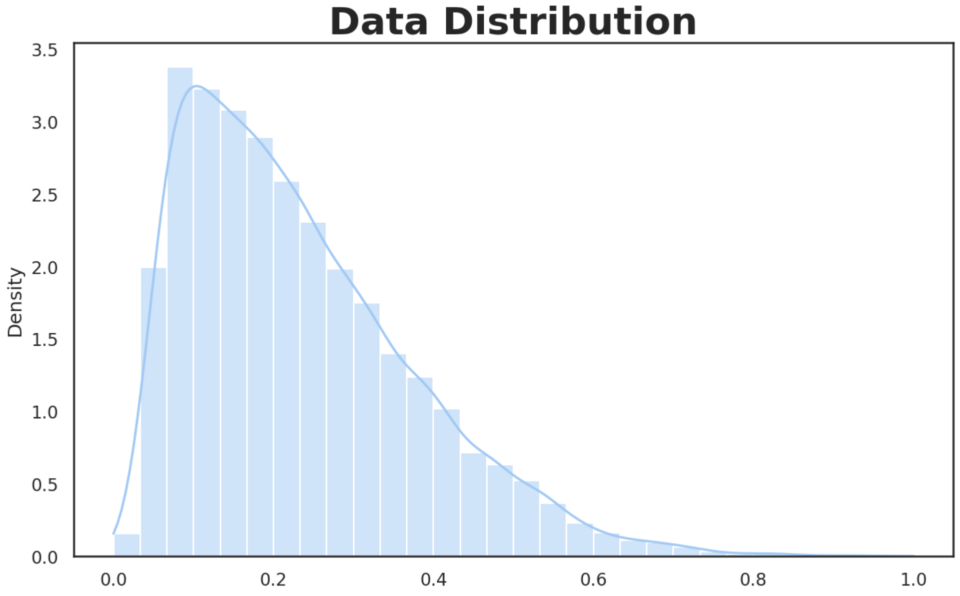
Add newly generated context to previous context

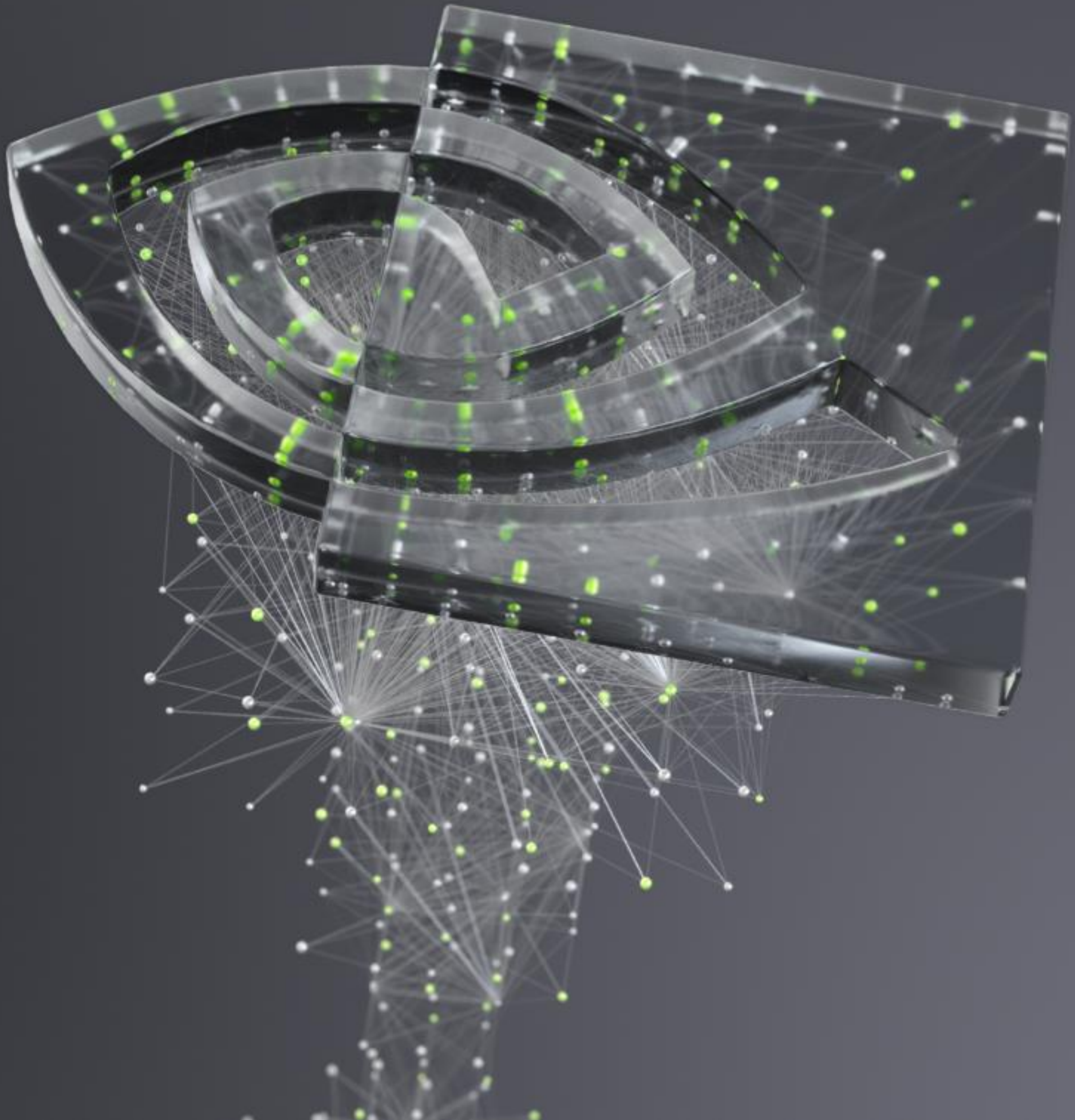


NVIDIA

MULTIPLE SAMPLING TO ESTIMATION DISTRIBUTION







EVALUATION

CREDIT CARD DATA

Real vs Synthetic

Real – 24M rows

user	card	amount	date	year	month	day	hour	minute	use chip	merchant name	merchant city	merchant state	zip	mcc	errors	is fraud
791	1	68.00	2018-01-02 09:10:00	2018	1	2	9	10	Swipe Transaction	12345536	New York	NY	10017	8005	<NA>	0
1572	0	572.42	2018-04-12 07:11:00	2018	4	12	7	11	Chip Transaction	49908535	Princeton	NJ	19406	5634	<NA>	0
2718	7	123.10	2019-01-04 10:14:00	2019	1	4	10	14	Chip Transaction	43211536	Beverly Hills	CA	90210	4800	<NA>	0
21	2	42.04	2020-06-23 11:18:00	2020	6	23	11	18	Swipe Transaction	65423006	Burke	VA	22015	5604	<NA>	0
1001	1	5000.00	2020-11-03 01:22:00	2020	11	3	1	22	Online Transaction	75434546	<NA>	<NA>	<NA>	1234	<NA>	1

Synthetic – 42M rows

user	card	amount	date	year	month	day	hour	minute	use chip	merchant name	merchant city	merchant state	zip	mcc	errors	is_fraud
1010	3	68.64	2019-07-22 12:43:00	2019	7	22	12	43	Chip Transaction	2027553650310142703	Boxford	MA	01921	5541	<NA>	0
142	0	2.21	2004-10-07 06:08:00	2004	10	7	6	8	Swipe Transaction	-6571010470072147219	Seattle	WA	98102	5499	<NA>	0
1037	1	24.32	2014-11-23 17:41:00	2014	11	23	17	41	Swipe Transaction	3959361429988996167	Tucson	AZ	85719	5912	<NA>	0
1734	0	29.60	2004-11-26 22:20:00	2004	11	26	22	20	Swipe Transaction	-4530600671233798827	Menlo Park	CA	94025	5812	<NA>	0
118	1	60.72	2018-11-16 21:53:00	2018	11	16	21	53	Chip Transaction	4751695835751691036	Anaheim	CA	92801	5814	<NA>	0

[Tabular Transformers for Modeling Multivariate Time Series (arxiv.org)]

EVALUATION FRAMEWORK TO MEASURE QUALITY OF GENERATED DATA

Tiered STOP-GO approach with example questions

Dataset Specific

1. **Coarse grained**
   a. Privacy: What % of data is a direct copy of the real data?
   b. What % of data is a self copy?
2. **Medium grained**
   a. Compare real column to synthetic column distributions
      - eg. Chi2 or Wasserstein
   b. Compare aggregate trends
3. **Fine grained**
   a. Compare joint distributions
   b. Privacy: Look for copied trajectories
      - eg. User in transacting with same merchants in order in both datasets. How long are these trajectories?
   c. Entity-level, Behavioral, Geographic, etc. trends followed.
      - eg. Big company is similar size in both datasets

COARSE GRAINED EVALUATION

- Privacy: What % of data is a direct copy of the real data?
  - 2 Rows

```
total = cudf.concat([real_df, synth_df])

copies = len(total) - len(total.drop_duplicates()) - (len(real_df) - len(real_df.drop_duplicates())) - (len(synth_df) - len(synth_df.drop_duplicates()))
```

Num_Copies = num_duplicated_rows(total) - num_duplicated_rows(real) – num_duplicated_rows(synthetic)
2 = 9129 - 66 - 9061

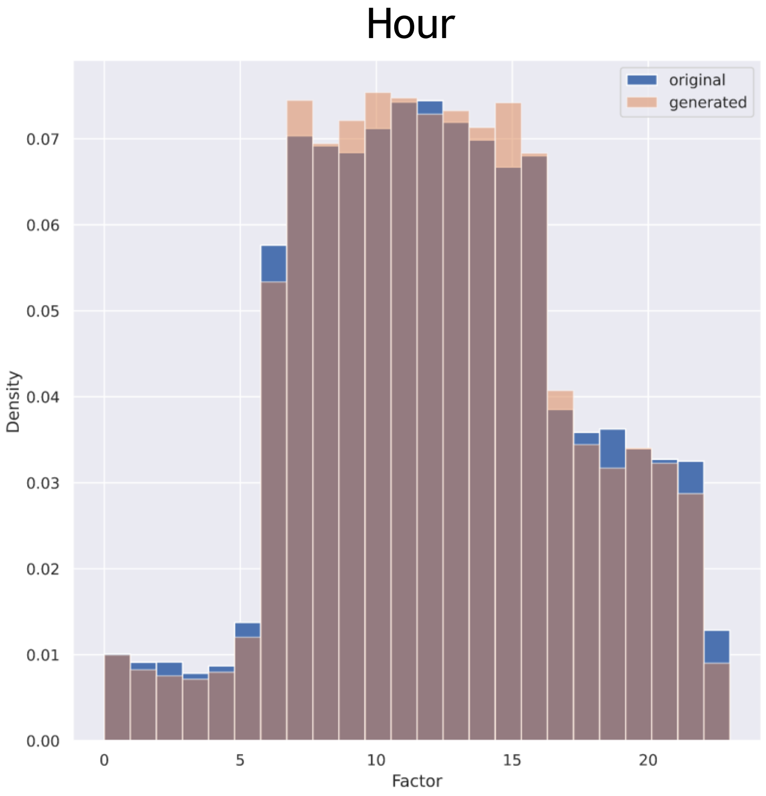
- What % of data is a self copy (duplicate rows)?
  - 0.02% in synthetic data

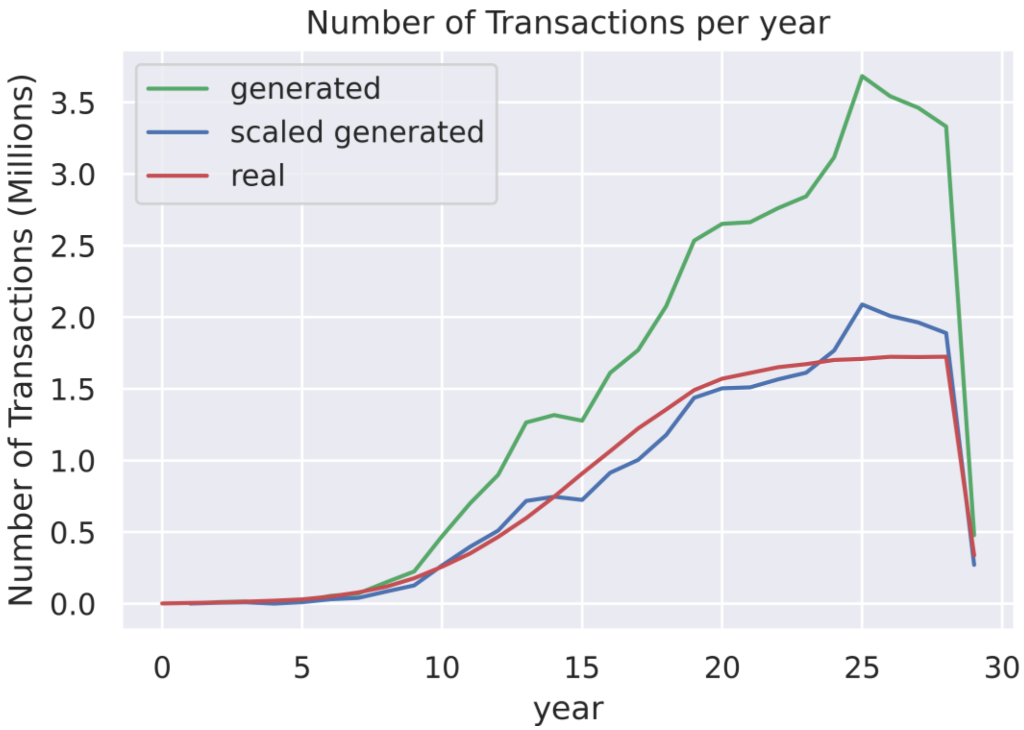
```
100*(len(synth_df) - len(synth_df.drop_duplicates()))/len(synth_df)
```

  - 0.0003% in the real data

MEDIUM GRAINED EVALUATION

- Compare real column to synthetic column distributions
  - eg. Chi2 or Wasserstein
- Compare aggregate trends

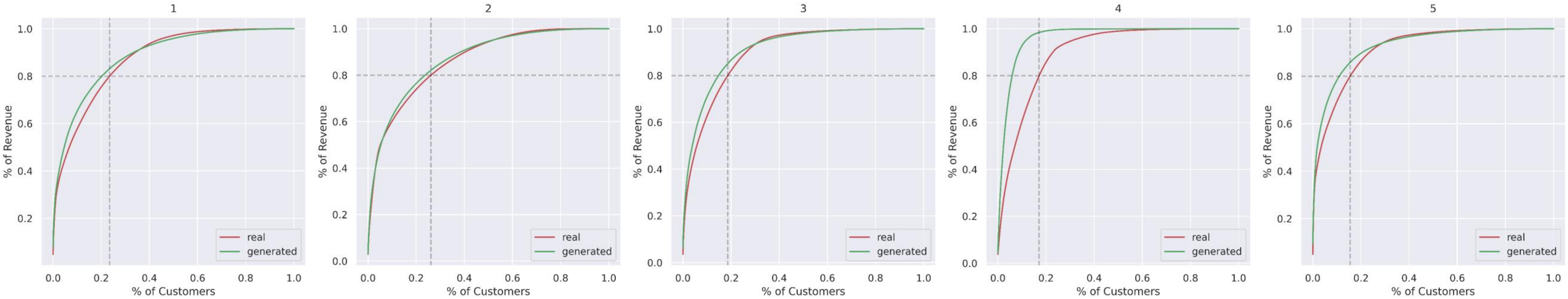




FINE GRAINED EVALUATION

- Compare joint distributions
- Privacy: Look for copied trajectories
  - eg. User in transacting with same merchants in order in both datasets. How long are these trajectories?
- Entity-level, Behavioral, Geographic, etc. trends followed.
  - eg. Big company is similar size in both datasets

merchant Id	Real txn counts	synthetic scaled txn counts	Ratio
59935	1130230	1076451	1.05
27092	1129061	999057	1.13
61195	1028485	1040211	0.99
39021	720615	633416	1.14
43293	687779	685414	1.00
14528	573491	644333	0.89



PLEASE RUN NOTEBOOKS

ETL/Preprocess → Model Training → Inference → Evaluation

Stop training notebooks

2_Training_Megatron.ipynb
2b_Tensorboard.ipynb

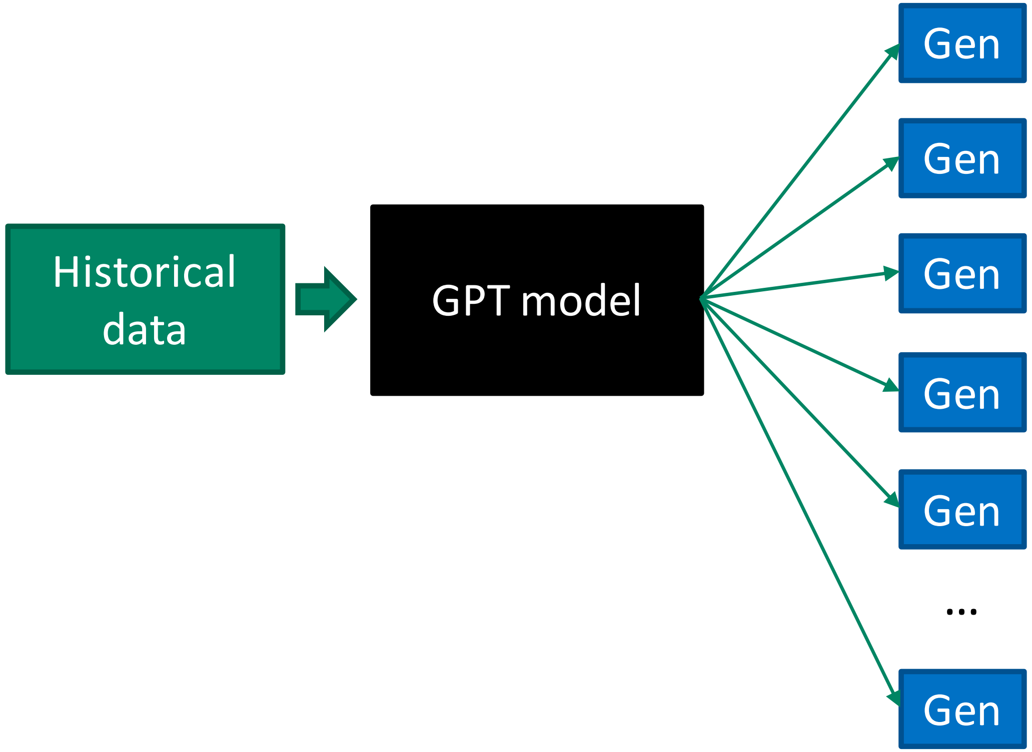
Start Inference and Eval notebooks

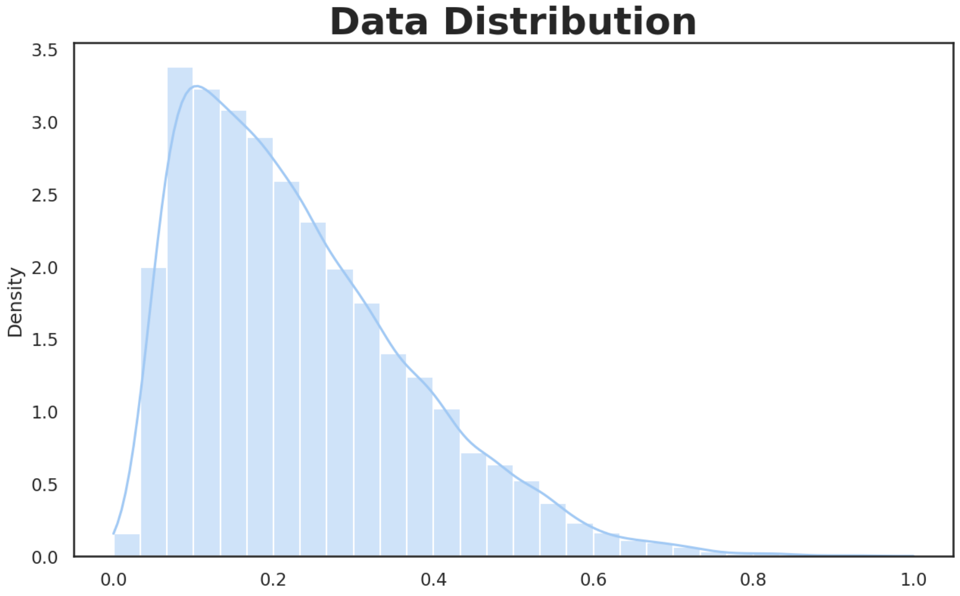
3a_Start_Inference_Server.ipynb
3b_Inference.ipynb
4_Evaluation.ipynb

Estimated time: 30 min

BONUS CASE STUDY: INFLATION
COHEN & STEERS

MULTIPLE SAMPLING TO ESTIMATION DISTRIBUTION





INFLATION

Understanding the trend of inflation particularly is important in identifying key turning points in the economy, central bank policy, and markets

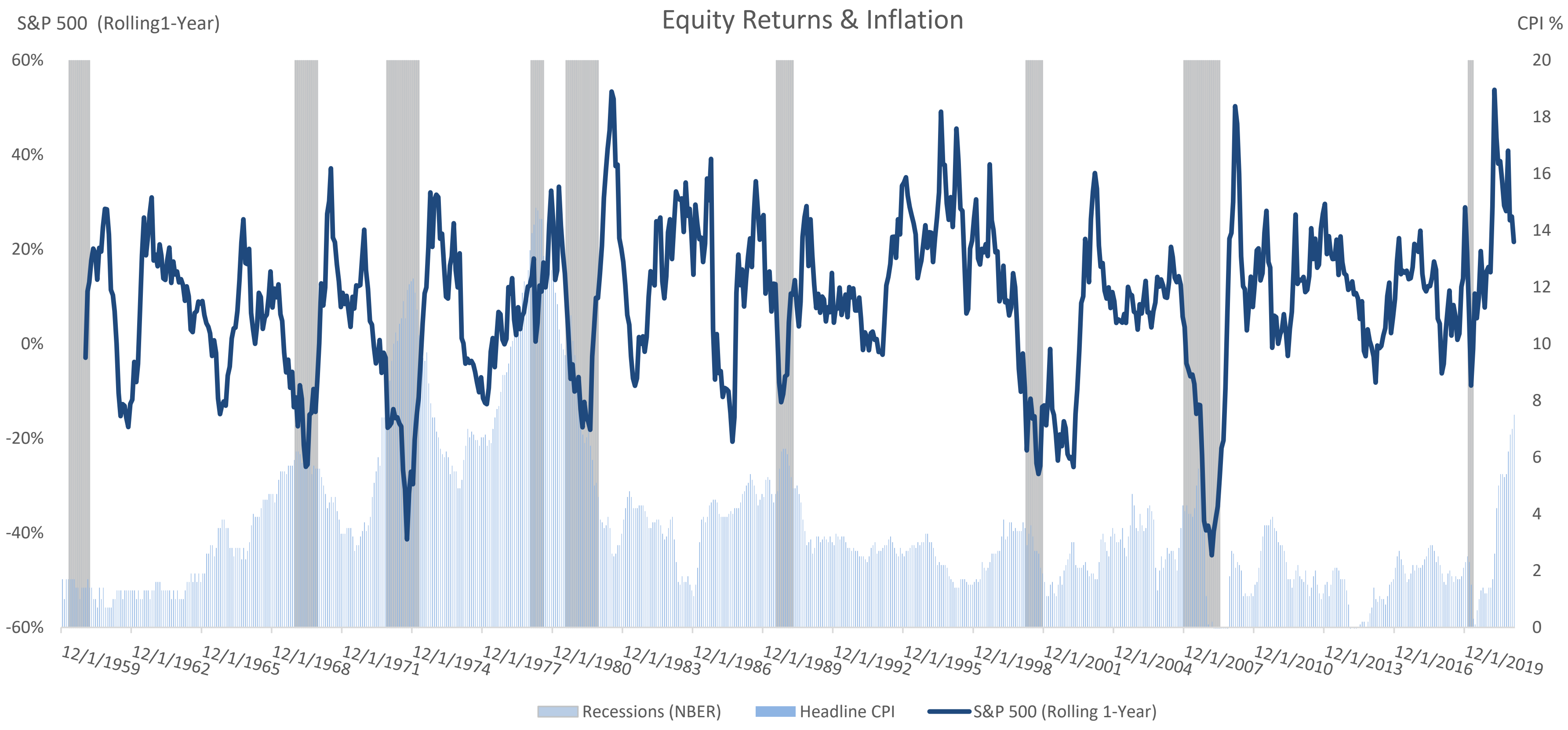


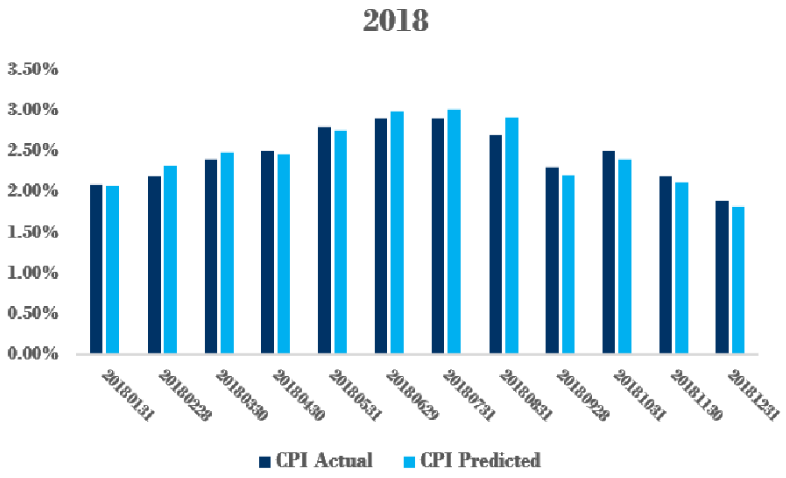
Figure 2 Source: Bloomberg, NBER)

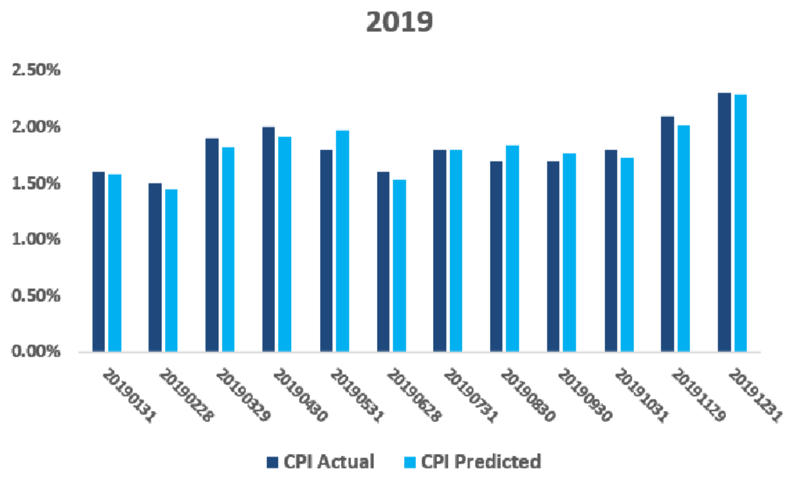
DATA INPUTS

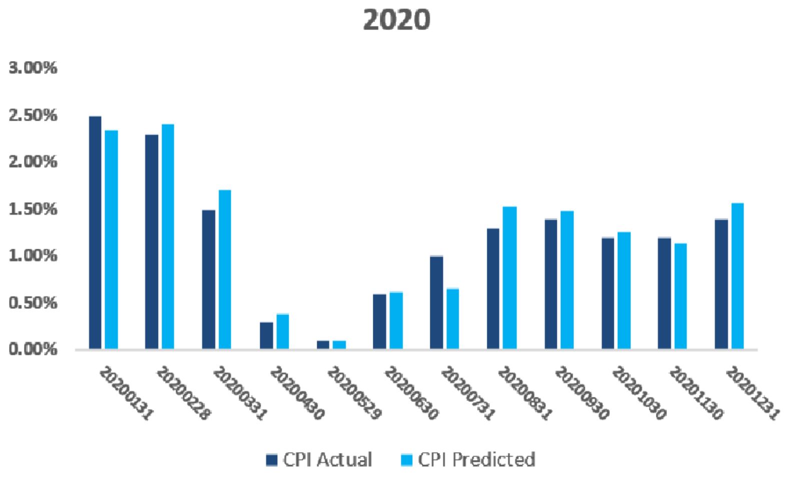
- Inputs include broad market and sector indexes, oil, credit spreads, term structure, money supply

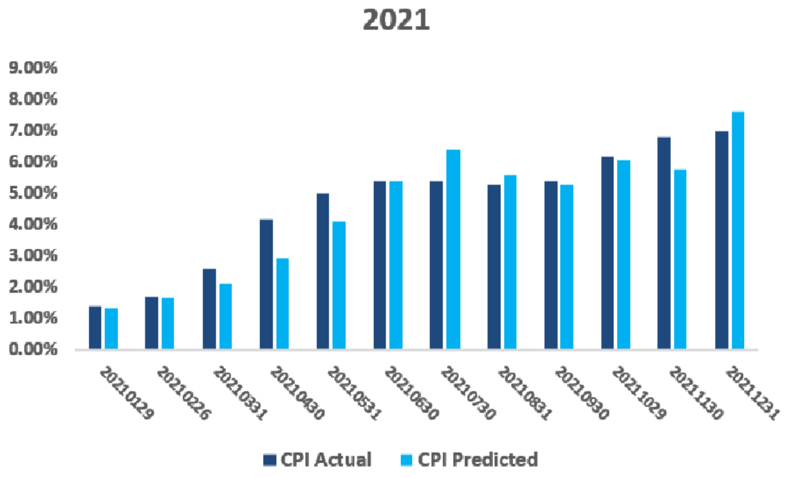
	S5INFT	S5FINL	S5INDU	S5CONS	S5RLST	S5UTIL	SPX	VIX	CPUPXYOY	USGG10YR	BCOM	USCRWTIC	USGG5YR	IBOXUMAE	GDP CHWG	USURTOT	M2	BSPGCPUS
Dec 29th 1989	64.7	76.14	75.11	62.33	-1E+09	103.76	353.4	-1E+09	4.4	7.935	93.149	21.82	7.832	-1E+09	9263	5.4	3152.5	-1E+09
Jan 1st 1990	64.7	76.14	75.11	62.33	-1E+09	103.76	353.4	-1E+09	4.4	7.935	93.149	21.82	7.832	-1E+09	9263	5.4	3152.5	-1E+09
Jan 2nd 1990	67.14	77.35	76.87	62.86	-1E+09	104.6	359.7	17.24	4.4	7.93	94.277	22.89	7.847	-1E+09	9263	5.4	3152.5	-1E+09
to	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
Dec 14th 2021	2969.58	644.4	870.72	780.03	306.8	351.31	4634	21.89	5	1.4411	95.984	70.73	1.2353	53.964	19469.4	4.2	21187	38.2
Dec 15th 2021	3051.33	646.42	878.47	789.26	311.3	357.21	4710	19.29	5	1.4565	95.518	70.87	1.2451	52.425	19469.4	4.2	21187	38.2
Dec 16th 2021	2963.95	654.23	878.92	793.61	312.6	358.96	4669	20.57	5	1.4106	97.035	72.38	1.1637	52.453	19469.4	4.2	21187	38.2
Dec 17th	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict
Dec 18th	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict
250 Days forward	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict	Predict

INFLATION INFERENCE 2018-2021

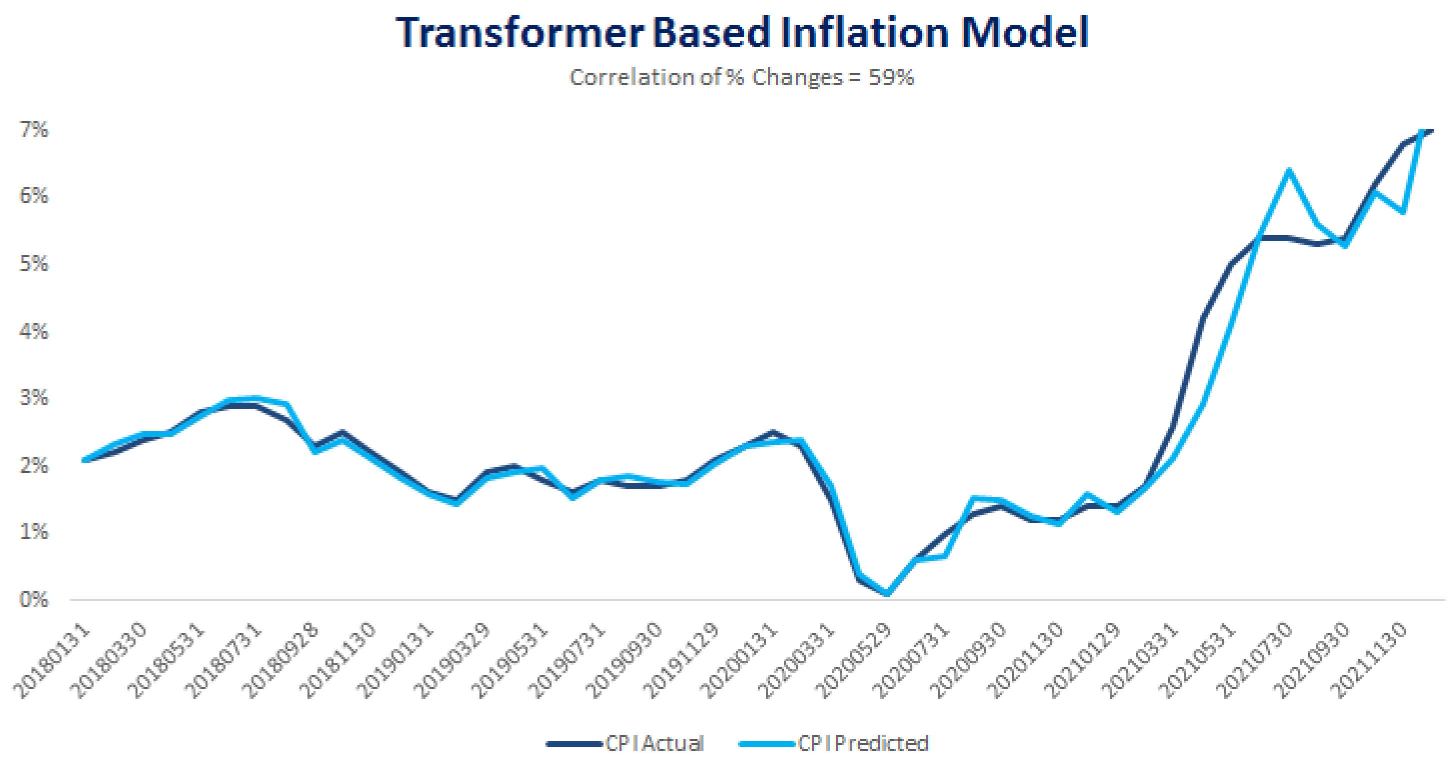






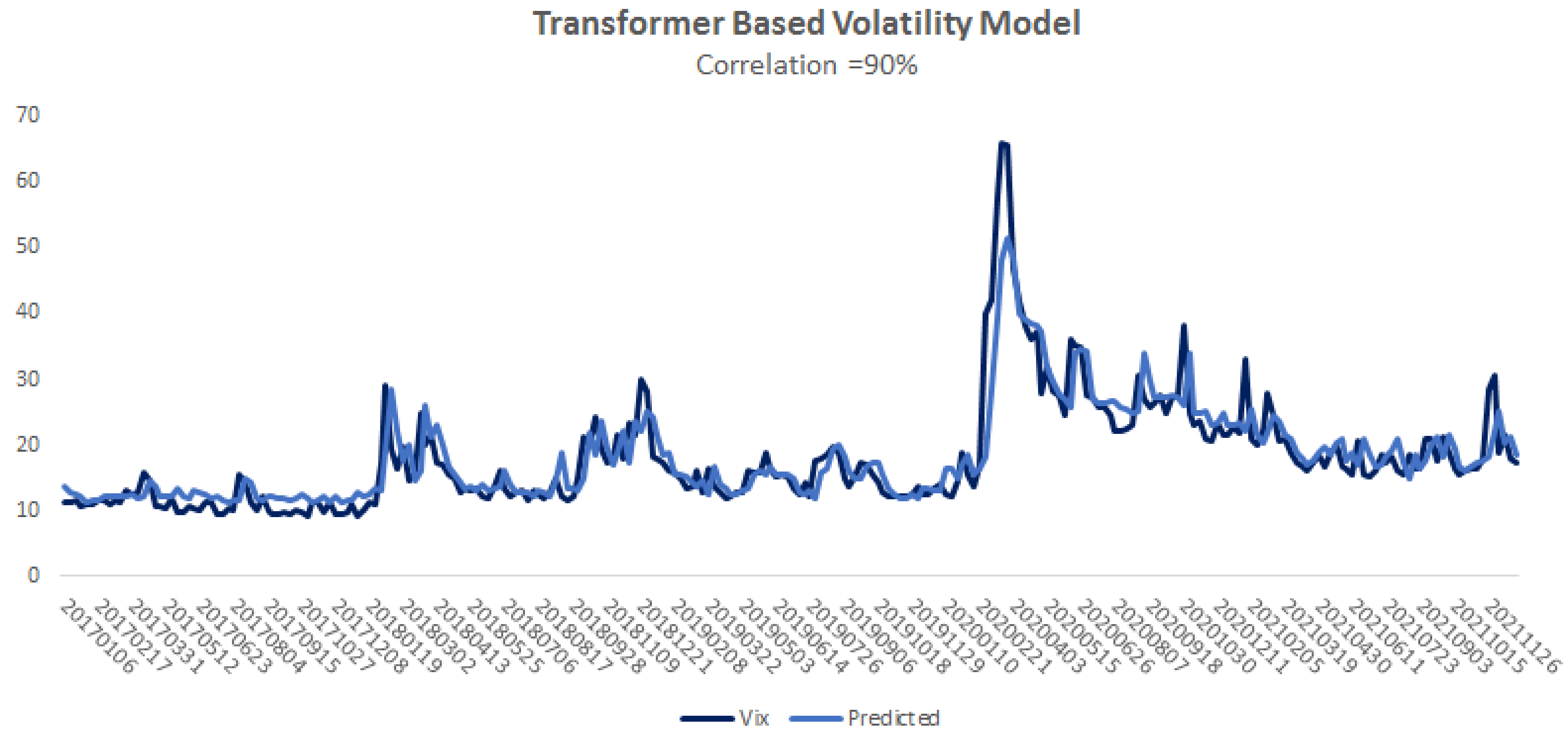


SUPERVISED LOSS: TRANSFORMER PREDICTIONS



NVIDIA.

VOLATILITY INFERENCE 2018-2021



SUMMARY AND NEXT STEPS

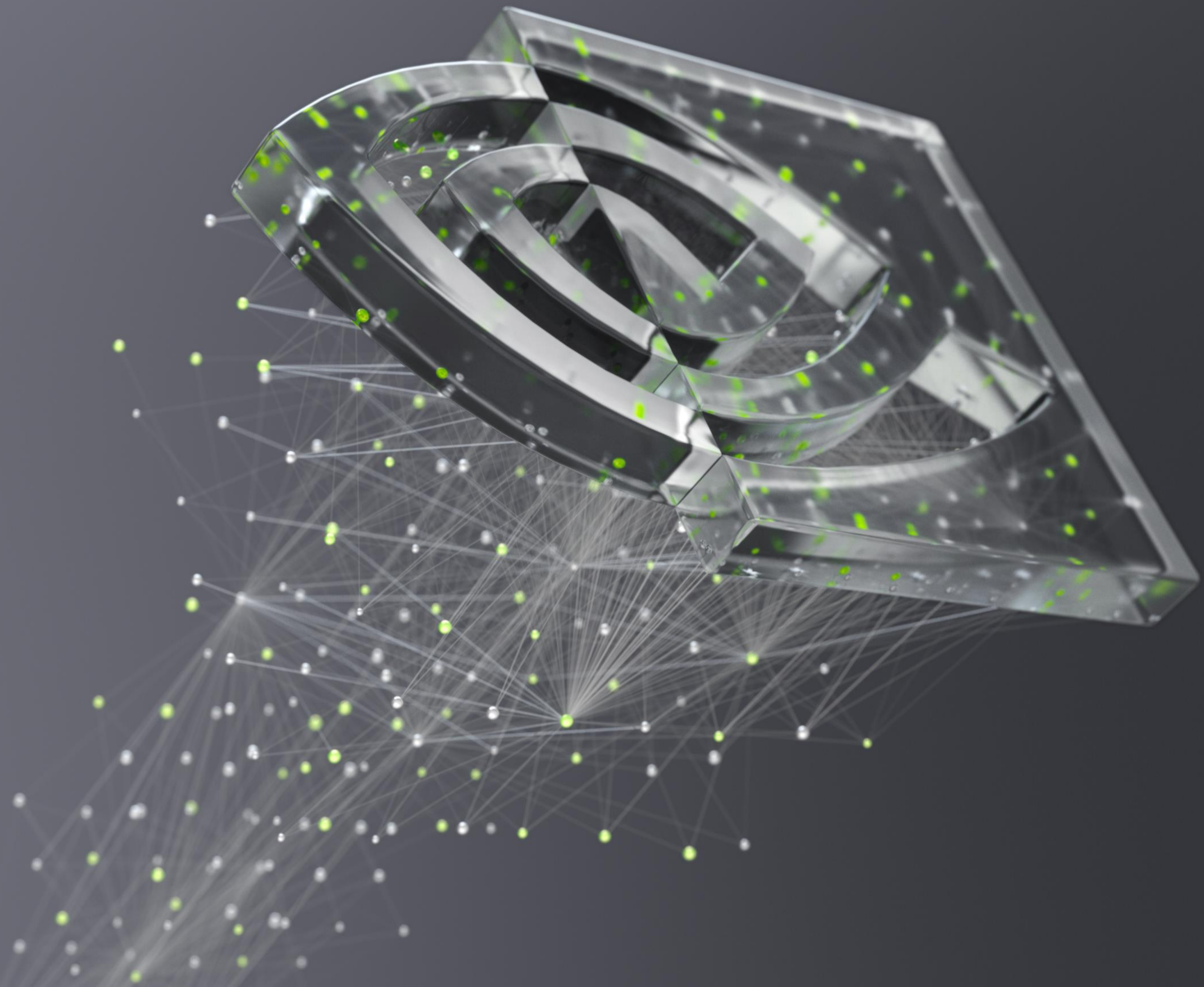
Here's what we covered:

- Tabular Synthetic Data Generation using GPT Transformer
- Tokenizing Tabular Data
- Training from scratch using NVIDIA Megatron
- Inference
- Evaluation

ETL/Preprocess → Model Training → Inference → Evaluation

Discussion & Next Steps: Use on your own data!

escoullos@nvidia.com



NVIDIA

APPENDIX

DATA SETS CHOSEN – LAST 31 YEARS – ~8400 ROWS DAILY CLOSING PRICE

SPX	S&P 500 INDEX
VIX	Chicago Board Options Exchange Volatility Index
CPUPXYOY	US CPI Urban Consumers Less Food & Energy YoY SA 1982=100
USGG10YR	US Generic Govt 10 Yr
BCOM	Bloomberg Commodity Index
USCRWTIC	US Crude Oil WTI Cushing OK Spot
USGG5YR	US Generic Govt 5 Yr
IBOXUMAE	MARKIT CDX.NA.IG.37 12/26
GDP CHWG	GDP US Chained 2012 Dollars SAAR
USURTOT	U-3 US Unemployment Rate Total in Labor Force Seasonally Adjusted
M2	Federal Reserve United States Money Supply M2 SA
BSPGCPUS	Federal Reserve Balance Sheet as a % of GDP
S5INFT	S&P 500 Information Technology Sector GICS Level 1 Index
S5FINL	S&P 500 Financials Sector GICS Level 1 Index
S5INDU	S&P 500 Industrials Sector GICS Level 1 Index
S5CONS	S&P 500 Consumer Staples Sector GICS Level 1 Index
S5RLST	S&P 500 Real Estate Sector GICS Level 1 Index