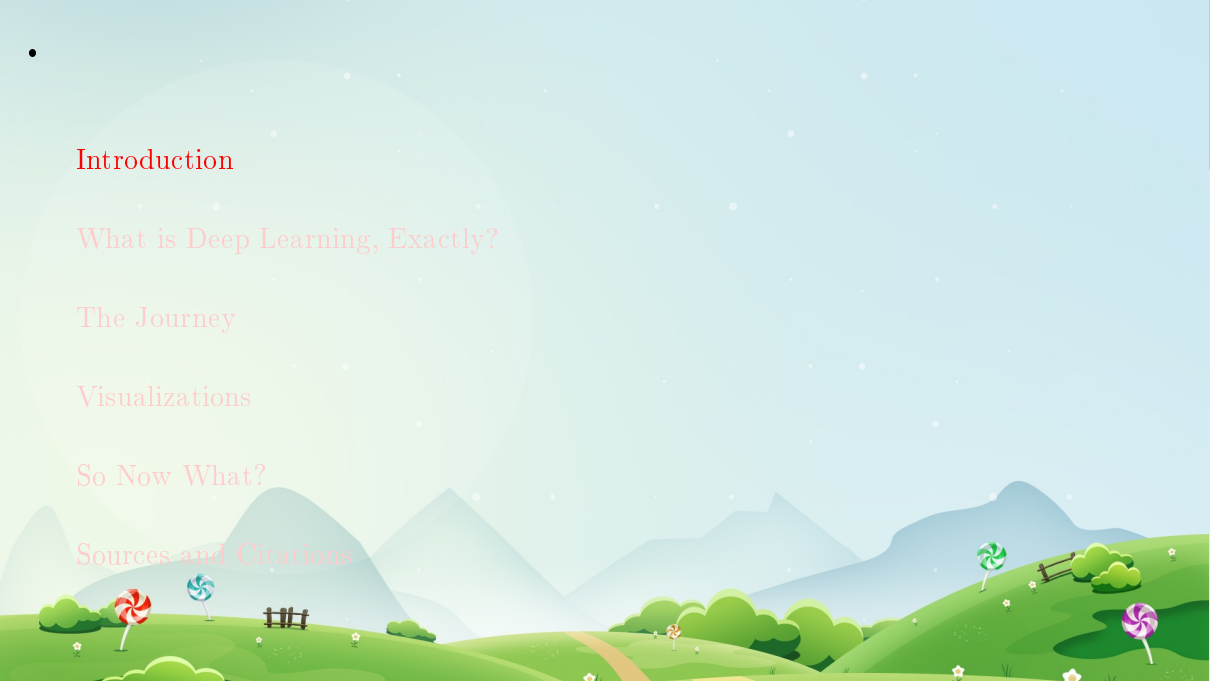


Learning About Deep Learning, and Maybe a Few Other Things

Franklin Diaz, Cosmic Voyager

2022-04-27

My presentation title is here. This talk is about my on-going journey into the world of graph theory and neural networks. This is a description of the learning process as well as the project and presenting some results.



Introduction

What is Deep Learning, Exactly?

The Journey

Visualizations

So Now What?

Sources and Citations

2022-04-27

└ Introduction

└ •

This first section is an Introduction to me and my project

Introduction

What is Deep Learning, Exactly?

The Journey

Visualizations

So Now What?

Sources and Citations

About Me

- I am a Security Consultant at Palo Alto Networks, cloud and automation for past 2+ years
- Did Data Eng/DevSecOps at Salesforce for 5 years.
- Been going to security conferences for a while.



2022-04-27

Introduction

- Here is a picture of me, but modified by a popular local artist.
- In my current role I get to work with the major cloud providers.
- In the past I was not a Data Scientist, but did some time on the Security Data Engineering team at Salesforce. This gave me a bit of a head start with data pipelines, directed acyclic graphs, and a few other things.

- I am a Security Consultant at Palo Alto Networks, cloud and automation for past 2+ years
- Did Data Eng/DevSecOps at Salesforce for 5 years.
- Been going to security conferences for a while.



The Project

- Realized that Terraform can output directed graphs.
- Had done a lot of work at Salesforce with directed graphs, data pipeline orchestration with AirFlow, etc. so I was somewhat familiar with the output I was seeing.
- The first question I had was, what can I do with these directed graphs?
- My hunch was I could “do some processing and analysis” of all this security infrastructure graph data and hoped that could lead to... predictions?

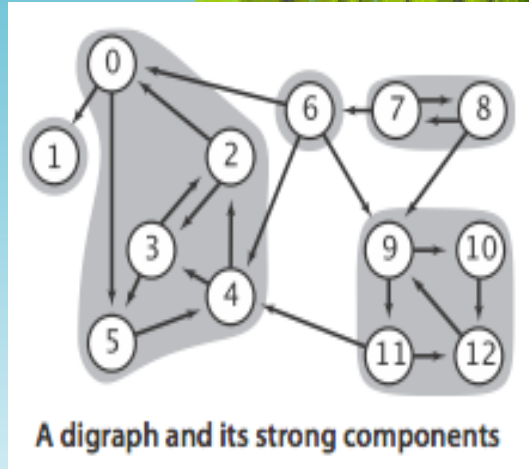
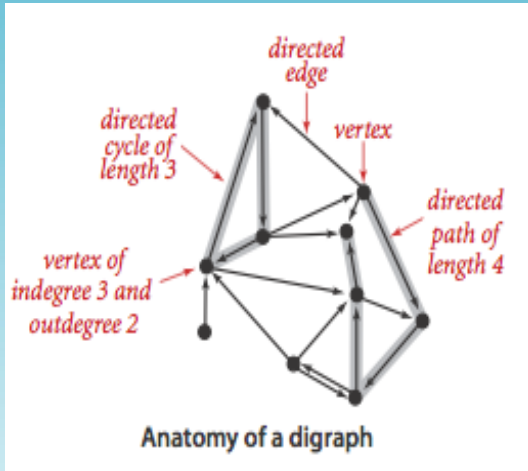
2022-04-27

Introduction

- Realized that Terraform can output directed graphs.
- Had done a lot of work at Salesforce with directed graphs, data pipeline orchestration with AirFlow, etc. so I was somewhat familiar with the output I was seeing.
- The first question I had was, what can I do with these directed graphs?
- My hunch was I could “do some processing and analysis” of all this security infrastructure graph data and hoped that could lead to... predictions?

- In case you are not familiar, Terraform is software that allows you to declare resources like network elements in public cloud providers.
- Had and still have this vague notion that if I had enough data I could find “outliers”. Maybe like a modernized version of a Pareto analysis?

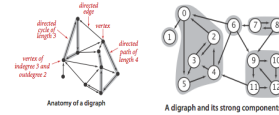
What's a DiGraph?



2022-04-27

Introduction

- The big takeaway here is the idea of “edges” and “nodes”
- Source: Algorithms, 4th Edition, by Robert Sedgewick and Kevin Wayne
- Wrath of Math!





Introduction

What is Deep Learning, Exactly?

The Journey

Visualizations

So Now What?

Sources and Citations

2022-04-27

└─ What is Deep Learning, Exactly?

└─ •

Introduction

What is Deep Learning, Exactly?

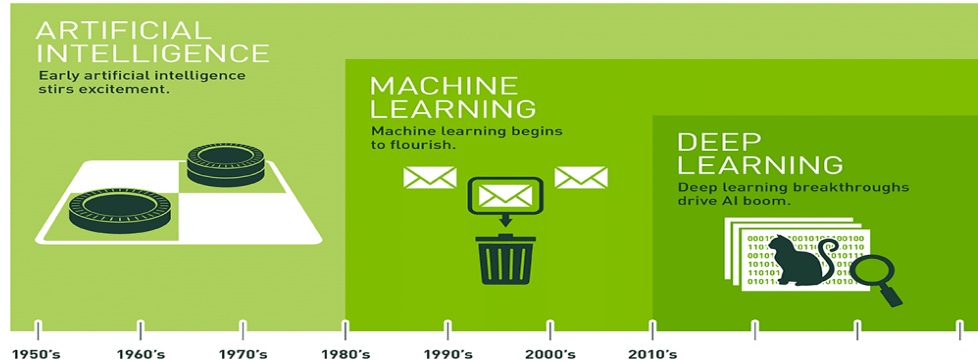
The Journey

Visualizations

So Now What?

Sources and Citations

The Rise of Deep Learning



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

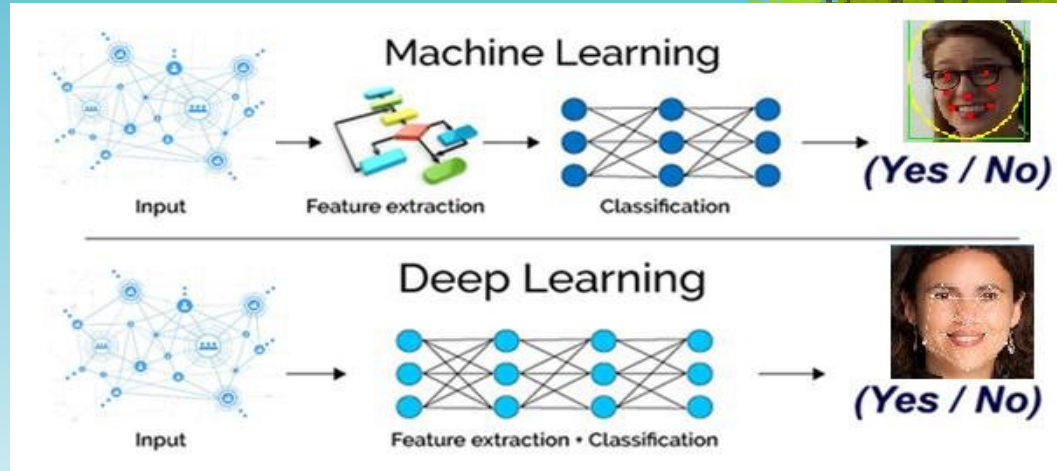
2022-04-27

What is Deep Learning, Exactly?

- GPUs have made it possible to expand accessibility to DL
- the CUDA toolkit from Nvidia has made things easier for researchers.

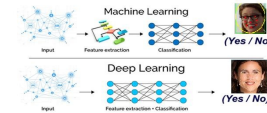


Quick Intro to a Giant Topic



2022-04-27

What is Deep Learning, Exactly?



- [image source/credit](#)
- ML feature extraction can be a huge undertaking, up to 80% of a project.
- DL attempts to automatically learn features that are most useful for a task from raw data.
- The nodes in a digraph are “neurons” or “units” in the DL/graph theory context.
- The neurons perform two steps. They calculate a “weighted sum” and pass the result through an “activation function” such as a rectifier activation function.
- These neurons or units that go through the rectifier function are called “RelUs” for short. Lot’s of descriptive info in this one term!
- Depth of the GNN is measured by the number of connected layers.
- DL needs very large data sets for accurate feature determination. Data sets with lots of features are known as “high density”.
- We humans interpret the features and output based on what we are trying to model.

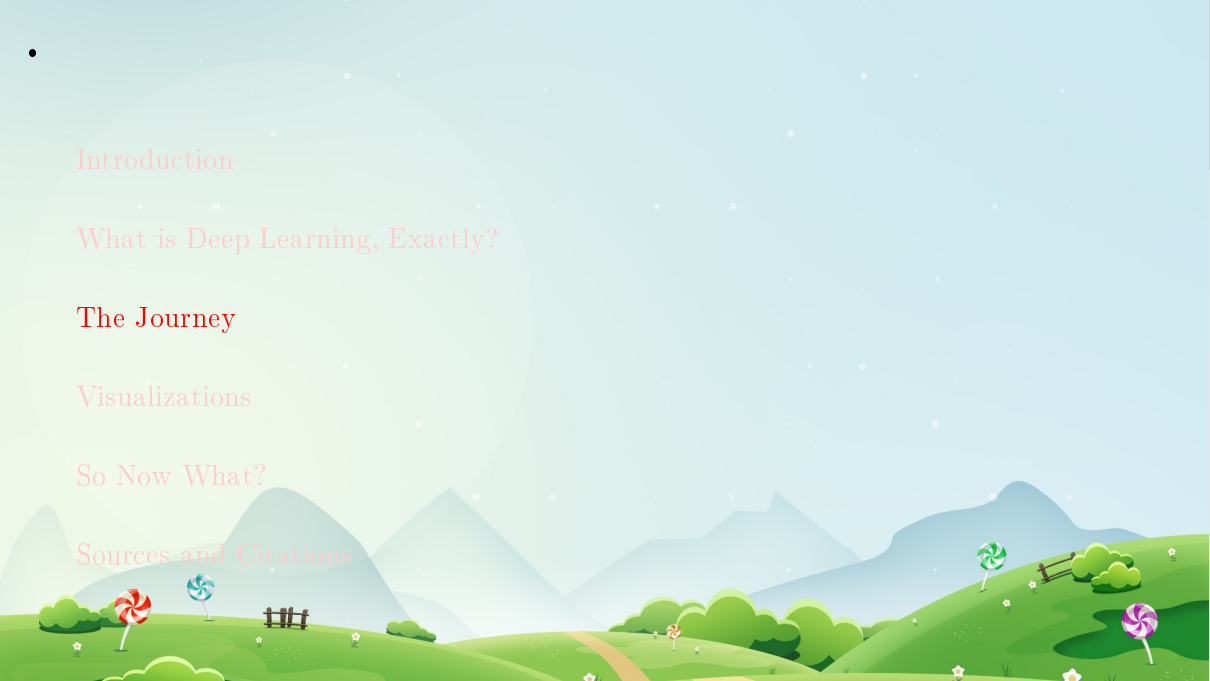
Amazing Tools Available



2022-04-27

└ What is Deep Learning, Exactly?

- Google Deep Learning Container Images
- Continuous Machine Learning (CML) Project
- Kaggle and shared Jupyter Notebooks



Introduction

What is Deep Learning, Exactly?

The Journey

Visualizations

So Now What?

Sources and Citations

2022-04-27

The Journey

•

Introduction

What is Deep Learning, Exactly?

The Journey

Visualizations

So Now What?

Sources and Citations

What Have I Gotten Myself Into?

- This is an example of a list.
- Important business information.



2022-04-27

└ The Journey

- This is an example of a list.
- Important business information.

Yak Shaving, Side Quests, Endless Rabbit Holes

- Makefiles and GNU Autotools
- NVIDIA Jetson Nano as cluster nodes
- SLURM cluster scheduler
- OpenMPI for parallel builds
- Docker and Containers
- k8s and Rancher k3s
- Data Version Control dvc.org
- Storing/accessing data in GCP buckets
- Continuous Machine Learning cml.dev
- Internal PyPI and Debian/Raspbian mirror (used too much bandwidth on home connection)

2022-04-27

└─ The Journey

- Makefiles and GNU Autotools
- NVIDIA Jetson Nano as cluster nodes
- SLURM cluster scheduler
- OpenMPI for parallel builds
- Docker and Containers
- k8s and Rancher k3s
- Data Version Control dvc.org
- Storing/accessing data in GCP buckets
- Continuous Machine Learning cml.dev
- Internal PyPI and Debian/Raspbian mirror (used too much bandwidth on home connection)

Dot Data Collection

A big barrier to entry was removed by the ability to output a Directed Graph from Terraform. [Click for video](#)

```
# Generate a PNG from Terraform
```

```
terraform graph | dot -Tpng > graph.png
```

```
# Generate vector graphic from Terraform
```

```
terraform graph | dot -Tsvg -o graph.svg
```

2022-04-27

└─ The Journey

A big barrier to entry was removed by the ability to output a Directed Graph from Terraform. [Click for video](#)

```
# Generate a PNG from Terraform  
terraform graph | dot -Tpng > graph.png
```

```
# Generate vector graphic from Terraform  
terraform graph | dot -Tsvg -o graph.svg
```



Python Data Collection

This became the basis for collection via Python.
[Click for video](#)

2022-04-27

└─ The Journey

This became the basis for collection via Python.
[Click for video](#)

Data Processing

github.com/devsecfranklin/model-html/pull/7

testing #7

devsecfranklin wants to merge 3 commits into main from dev5-20211118

github-actions bot commented on 39f7988 on Nov 19, 2021

Creating data diretory.
Downloading the dataset.
Count: ip 10285
transport 10285
port 10285
server 10285
dtype: int64

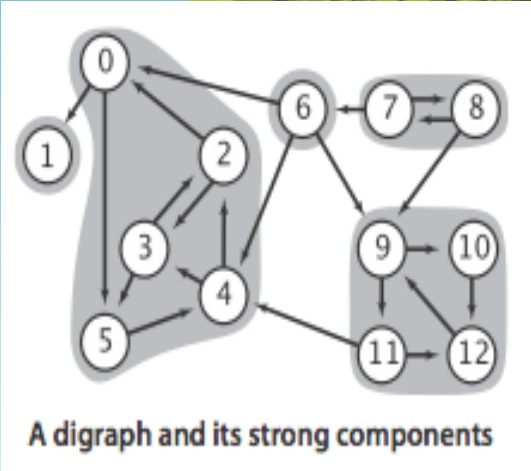
Data sample:

	ip	transport	port	server
0	160.153.63.6	tcp	443	[Apache]
1	35.197.16.105	tcp	80	[Apache]
2	199.19.210.59	tcp	443	[Apache/2.4.10', 'Unix', 'OpenSSL/1.0.1]]
3	12.130.77.39	tcp	443	[Apache]
4	129.121.5.237	tcp	443	[Apache]

generating graph object in networkx

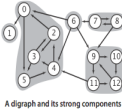
github-actions bot commented on d14cde9 on Nov 19, 2021

Creating data diretory.
Downloading the dataset.
Count: ip 10285



2022-04-27

The Journey





Data Collection Container

2022-04-27

└ The Journey

Data Storage - Google Cloud

Data storage with GCP because it's (relatively) easy.



2022-04-27

└ The Journey

Data storage with GCP because it's (relatively) easy.

Data Storage - Kaggle

Data storage and tagging using DVC



2022-04-27

└ The Journey

Data storage and tagging using DVC

Data Storage - DVC

Data storage and tagging using DVC



2022-04-27

└ The Journey

Data storage and tagging using DVC

Data Tagging

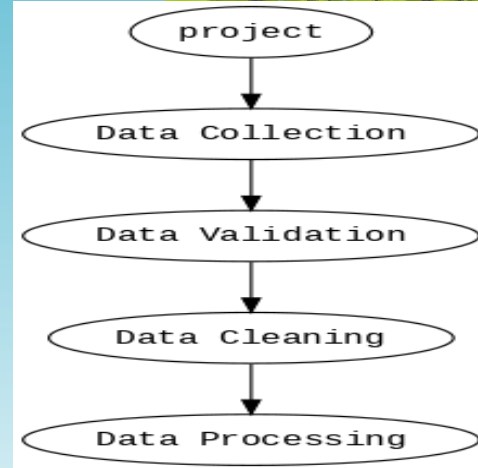


2022-04-27

└ The Journey

Data Pipeline

- The Data Pipeline is a set of processes that move and transform data from various sources to a destination where new value can be derived.
- The DP is the foundation of analytics, reporting, and machine learning capabilities.



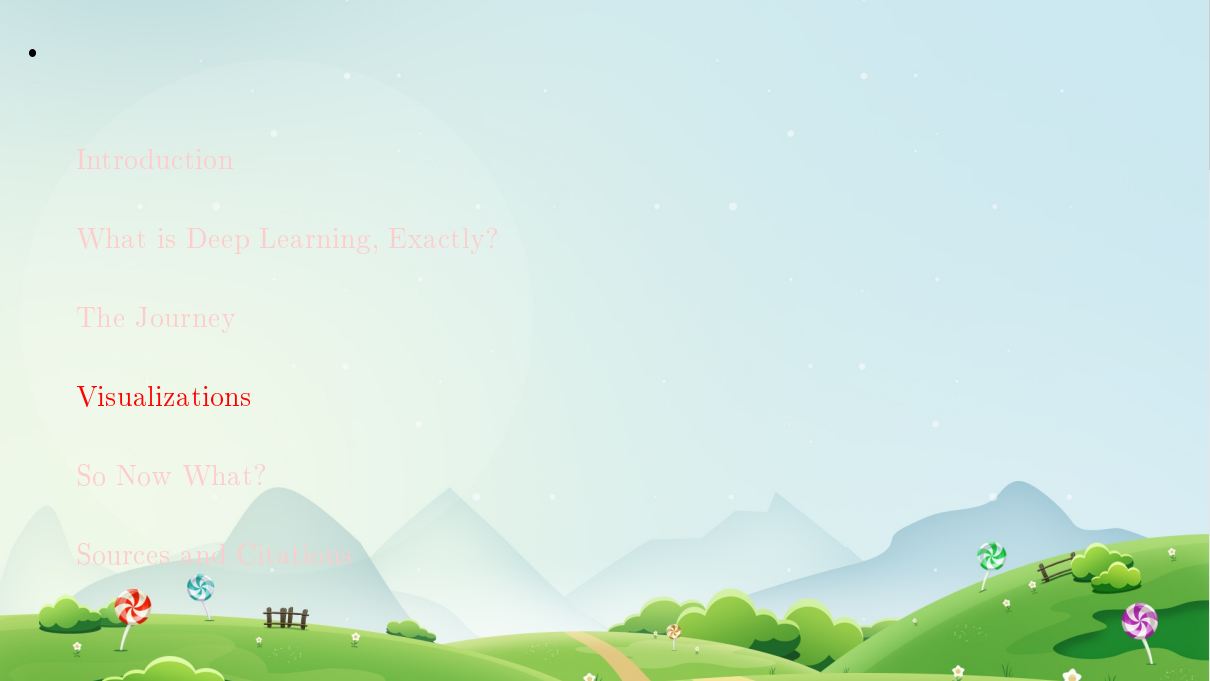
2022-04-27

The Journey

- Source: Data Pipelines pocket reference p1-2

- The Data Pipeline is a set of processes that move and transform data from various sources to a destination where new value can be derived.
- The DP is the foundation of analytics, reporting, and machine learning capabilities.





Introduction

What is Deep Learning, Exactly?

The Journey

Visualizations

So Now What?

Sources and Citations

2022-04-27

└ **Visualizations**

└ •

Introduction

What is Deep Learning, Exactly?

The Journey

Visualizations

So Now What?

Sources and Citations

2022-04-27

└ Visualizations

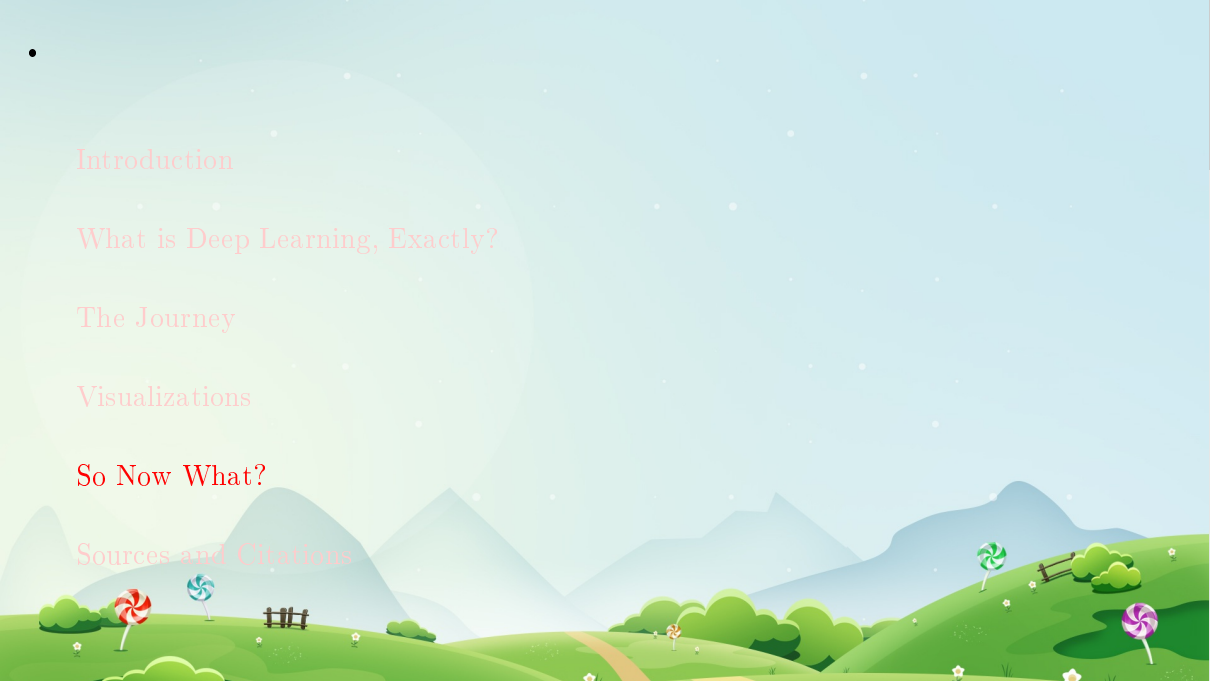
- This is the first thing I saw when I started converting the data.
- Was excited here since I was able to change the color of the nodes.
- Obviously this is not yet a usable result.

Gephi



2022-04-27

└ Visualizations



Introduction

What is Deep Learning, Exactly?

The Journey

Visualizations

So Now What?

Sources and Citations

2022-04-27

└─ So Now What?

└─ •

Introduction

What is Deep Learning, Exactly?

The Journey

Visualizations

So Now What?

Sources and Citations

Useful Intermediate Results

- Standardizing my data collection on JSON.
- Made some super cool functions for parsing nested JSON.
- Importing JSON to Pandas dataframes.

2022-04-27

└ So Now What?

- Tabular data in Pandas can be output in all kinds of formats.
- Pandas data frames can be the input for other Machine Learning tools and frameworks.

- Standardizing my data collection on JSON .
- Made some super cool functions for parsing nested JSON .
- Importing JSON to Pandas dataframes .

Next Steps for this Project

- See if I can get the training to use my personal GPU/TPU

2022-04-27

└ So Now What?

- Most of this work is relegated to my “free” time.
- Have to spend my days helping people with the cloud.

• See if I can get the training to use my personal GPU/TPU

If Time, Staff, and Money were no Object

- Dreams and stretch goals.

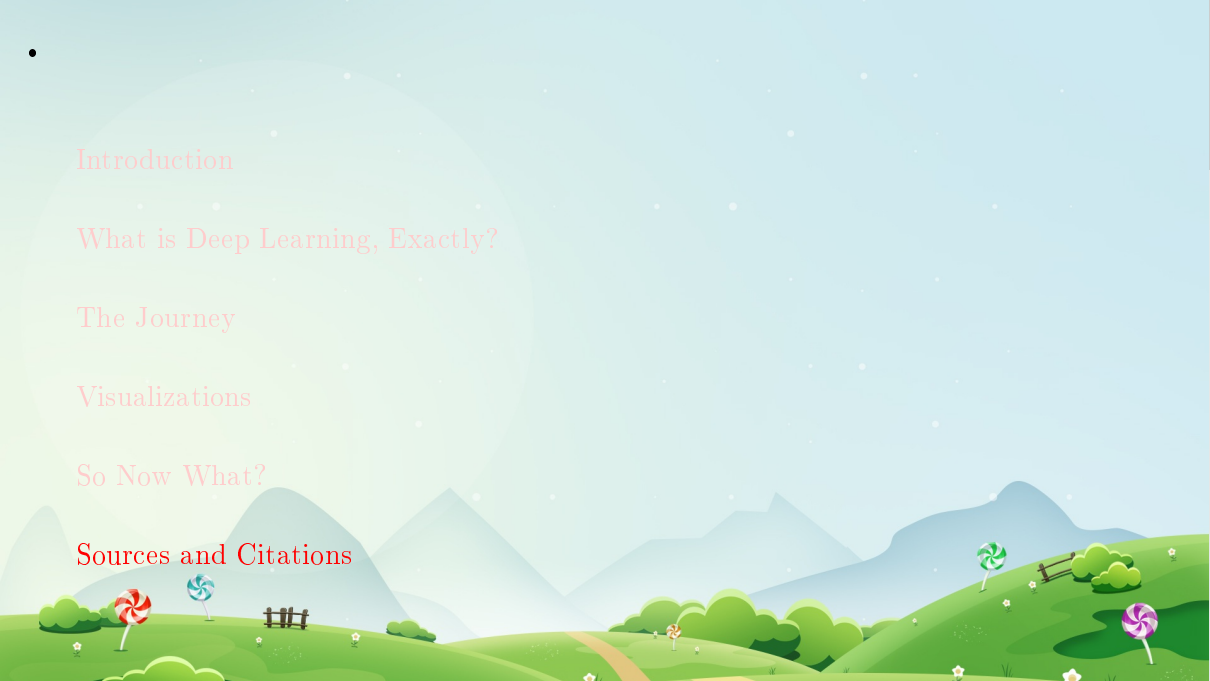


2022-04-27

└ So Now What?

- this is the “dreams and stretch goals” slide.
- Most of this work is relegated to my “free” time.
- Have to spend my days helping people with the cloud.

• Dreams and stretch goals.



Introduction

What is Deep Learning, Exactly?

The Journey

Visualizations

So Now What?

Sources and Citations

2022-04-27

Sources and Citations

Introduction

What is Deep Learning, Exactly?

The Journey

Visualizations

So Now What?

Sources and Citations

Sources and Citations

- This is an example of a list.
- Important business information.



2022-04-27

Sources and Citations

- point 1
- point 2

• This is an example of a list.

• Important business information.