

# Data Engineering

---

Blake Vente



# Constraints

- free hosting

# Constraints

- free hosting
- no fine print

# Constraints

- free hosting
- no fine print
- familiar

# Constraints

- free hosting
- no fine print
- familiar
- easy to maintain

# Constraints

- free hosting
- no fine print
- familiar
- easy to maintain
- mobile friendly

# Constraints

- free hosting
- no fine print
- familiar
- easy to maintain
- mobile friendly
- Screen-Reader Friendly (accessibility)



## **Github-centered Workflow**

- GitHub Pages

## **GitHub Markdown**

## **Github-centered Workflow**

- GitHub Pages
- GitHub Markdown

## **GitHub Markdown**

## **Github-centered Workflow**

- GitHub Pages
- GitHub Markdown
- Open Source

## **GitHub Markdown**

## **Github-centered Workflow**

- GitHub Pages
- GitHub Markdown
- Open Source

## **GitHub Markdown**

- superset of HTML

## **Github-centered Workflow**

- GitHub Pages
- GitHub Markdown
- Open Source

## **GitHub Markdown**

- superset of HTML
- reduces our maintenance complexity

- Losslessly convert between HTML and Markdown using Pandoc

- `_posts/labs/` stores the labs

- `_posts/labs/` stores the labs
- `_posts/readings/` stores the readings



- `_posts/labs/` stores the labs
- `_posts/readings/` stores the readings
- etc...

- `_posts/labs/` stores the labs
- `_posts/readings/` stores the readings
- etc...
- Push to master...

- `_posts/labs/` stores the labs
- `_posts/readings/` stores the readings
- `etc...`
- Push to master...
- Everything autopopulates in 5 minutes or less.

## Upgrading

- Ported model to Python 3<sup>1</sup>

---

<sup>1</sup><https://github.com/untrix/im2latex>

## Upgrading

- Ported model to Python 3<sup>1</sup>
- Future plans to upgrade Tensorflow to v2

---

<sup>1</sup><https://github.com/untrix/im2latex>

## Upgrading

- Ported model to Python 3<sup>1</sup>
- Future plans to upgrade Tensorflow to v2
- Deal with deprecated dependencies

---

<sup>1</sup><https://github.com/untrix/im2latex>

# Progress so far

## Training model

- Model has been deployed on a Google-Cloud GPU learner and locally for consistency

## Following Upgrades to Model

# Progress so far

## Training model

- Model has been deployed on a Google-Cloud GPU learner and locally for consistency
- Trains without problems - all log files show promising results

## Following Upgrades to Model



# Progress so far

## Training model

- Model has been deployed on a Google-Cloud GPU learner and locally for consistency
- Trains without problems - all log files show promising results
- Currently has problems restoring from checkpoint

## Following Upgrades to Model

# Progress so far

## Training model

- Model has been deployed on a Google-Cloud GPU learner and locally for consistency
- Trains without problems - all log files show promising results
- Currently has problems restoring from checkpoint

## Following Upgrades to Model

- Be able to tweak the attention model and customizable hyperparameters of the model

# Progress so far

## Training model

- Model has been deployed on a Google-Cloud GPU learner and locally for consistency
- Trains without problems - all log files show promising results
- Currently has problems restoring from checkpoint

## Following Upgrades to Model

- Be able to tweak the attention model and customizable hyperparameters of the model
- Upgrade preprocessing scripts to work on Blake's generated data for consistency

# Progress so far

## Training model

- Model has been deployed on a Google-Cloud GPU learner and locally for consistency
- Trains without problems - all log files show promising results
- Currently has problems restoring from checkpoint

## Following Upgrades to Model

- Be able to tweak the attention model and customizable hyperparameters of the model
- Upgrade preprocessing scripts to work on Blake's generated data for consistency
- Experiment with running the model on various data categories (i.e Matrices, Equations, Piecewise Functions)

- Setting up a Jekyll site with GitHub Pages<sup>2</sup>

---

<sup>2</sup><https://jekyllrb.com/docs/github-pages/>