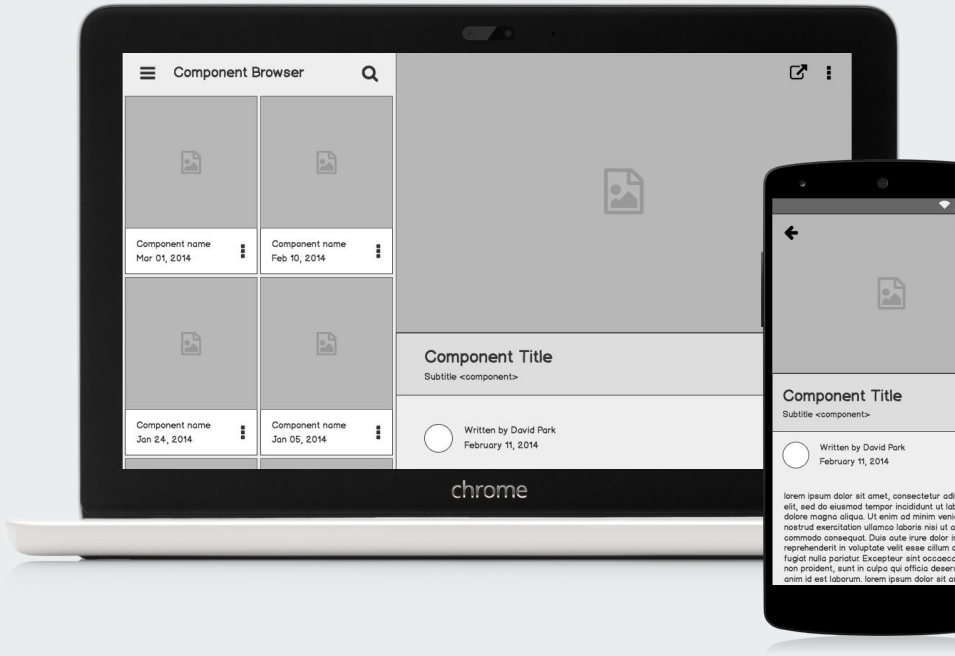
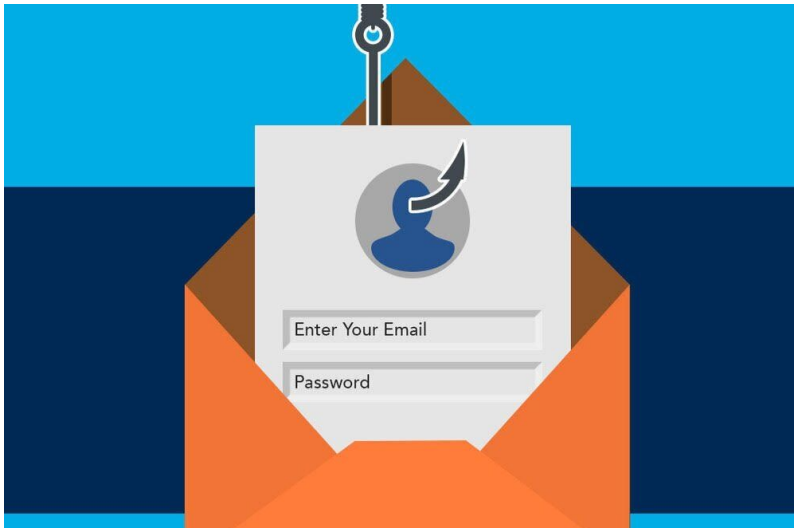
Fraudulent Job Postings Detection

Meilin Li, Yunhe Jia
2022/06/30



Goal

- Fraudulent postings are annoying and time-wasting in the job search process.
- Result in personal information leakage through application.
- Build NLP model to automatically detect fake job postings based on its content…



Data

- **Data Source:** Kaggle (https://www.kaggle.com/datasets/shivamb/real-or-fake-fake-jobposting-prediction)
- Number of postings: ~18k; Fake rate: ~5%
- Description + Requirements + Benefits => Raw Text
- Train-Valid-Test split: 70%-15%-15%

What We Tried

- Recurrent neural network (RNN)
- Convolutional neural network(CNN)
- Pre-trained BERT Model

Performance Comparison(Test Data)

	Accuracy	Precision	Recall	Training Time(10 epochs)
RNN	0.98	0.99	0.82	~13 mins
CNN	0.97	0.99	0.76	~9 mins
BERT	0.97	0.95	0.74	>1 day (2h on GPU)

What Went Well

- We successfully trained three different models to perform the task and it’s a good practice for what we learned on class
- The model is good at identifying fake job postings

What Went Wrong (Future Improvement)

- We used an imbalance dataset but no special dealing with it
- Takes a lot of time training locally (could use collab to train from start)
- Improve and wrap up the model to integrate in email/google extension.

Thank you!