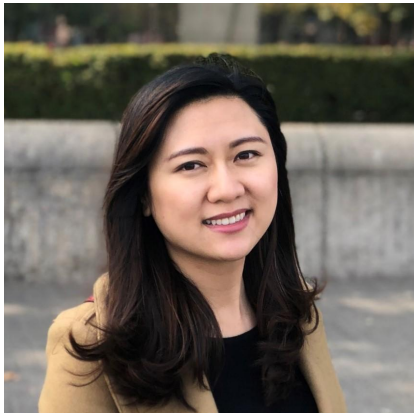
DD.ai

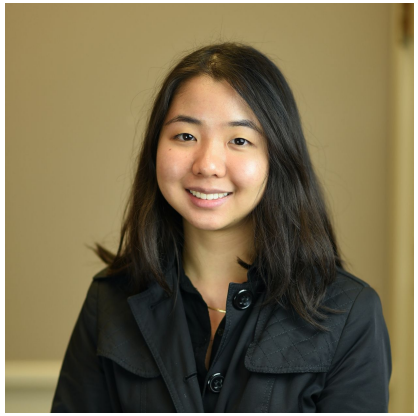
A Platform for Drug-Drug Interaction Prediction

Our Team



Mai La

Data Scientist
Data Engineer | MLOps



Grace Lee

Data Scientist
Product Manager | UI

Radhika Mardikar

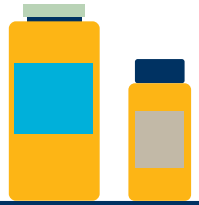
Data Scientist
ML Engineer | UI

01
Impact
Opportunity
Target Users

02
Product
Quick Demo

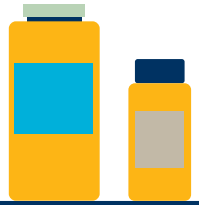
03
Approach
Details on data
Scoping
ML set up

04
Results
Evaluation
Future Work
Conclusion



01

Impact



Drug - Drug interactions (DDI) can have severe effects

- **52%** of adverse drug effects are **preventable**[1]
- Account for **18.3%** of all adverse drug effects[2]
- **22.4%** of adults in the US use **5+** medications[3]

[1] Source: Hakkarainen et al., 2012

[2] Source: Jiang et al., 2022

[3] Source: Hales et al., 2019

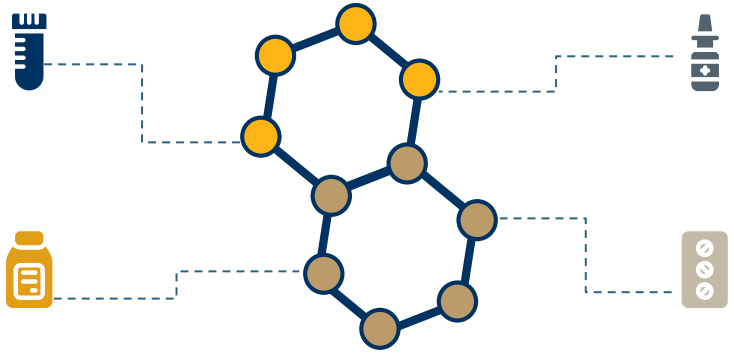
Drug Interaction Testing is Expensive

$60 Billion

2023 Global Spending[4]

60% increase

Drug approvals 2010-2019[6]



$133 Billion

2032 Global Spending[5]

4-5 Years

For testing full panel and reporting[7]

[4] Source: Precedence Research
[6] Source: CBO

[5] Source: Precedence Research
[7] Source: User Interviews

User Interviews

Pharmacist - Private

- Manually find DDI

Supplement Co CEO

- Report for FDA
- Lots of testing

Org Chem PhD

- Chemical Formula
- Mechanism of Action

Pharmacist - CVS

- Have built in database for searching

DD.ai

Optimize DDI testing and assist drug companies to bring new drugs to market

Input:

- SMILES (Simplified Molecular-Input Line-Entry System) Drug structure

```
COC1=CC2=C(C=C1)N=C(N2)S(=O)CC1=NC=C(C)C(OC)=C1C
```

- Mechanism of Action Pathway

```
This drug stops gastric acid secretion by selective inhibition of the H+/K+ ATPase enzyme system...
```

Output:

- Prediction Probability of Positive Interactions
- Breakdown of supplement vs medication
- Downloadable table with drug list and probabilities

02
Product

Demo

Product Feedback

Dr. Han Dang, PharmD
Pharmacy Manager

Impressive that AI could extract DDI for new drug development.

Mai Do, RPh. , MBA
Pharmacy Manager, CVS

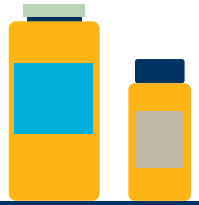
I was amazed by how fast this tool can help define possible interactions between a new drug with hundreds of available drugs at a very high level of accuracy. This definitely help drug discovery much faster and more efficient with a lower cost. Traditionally, the process of evaluating drug interactions might take months or years of experiments in vitro or in vivo. With DD.ai, it's less than 1 minute. Simply amazing!!!

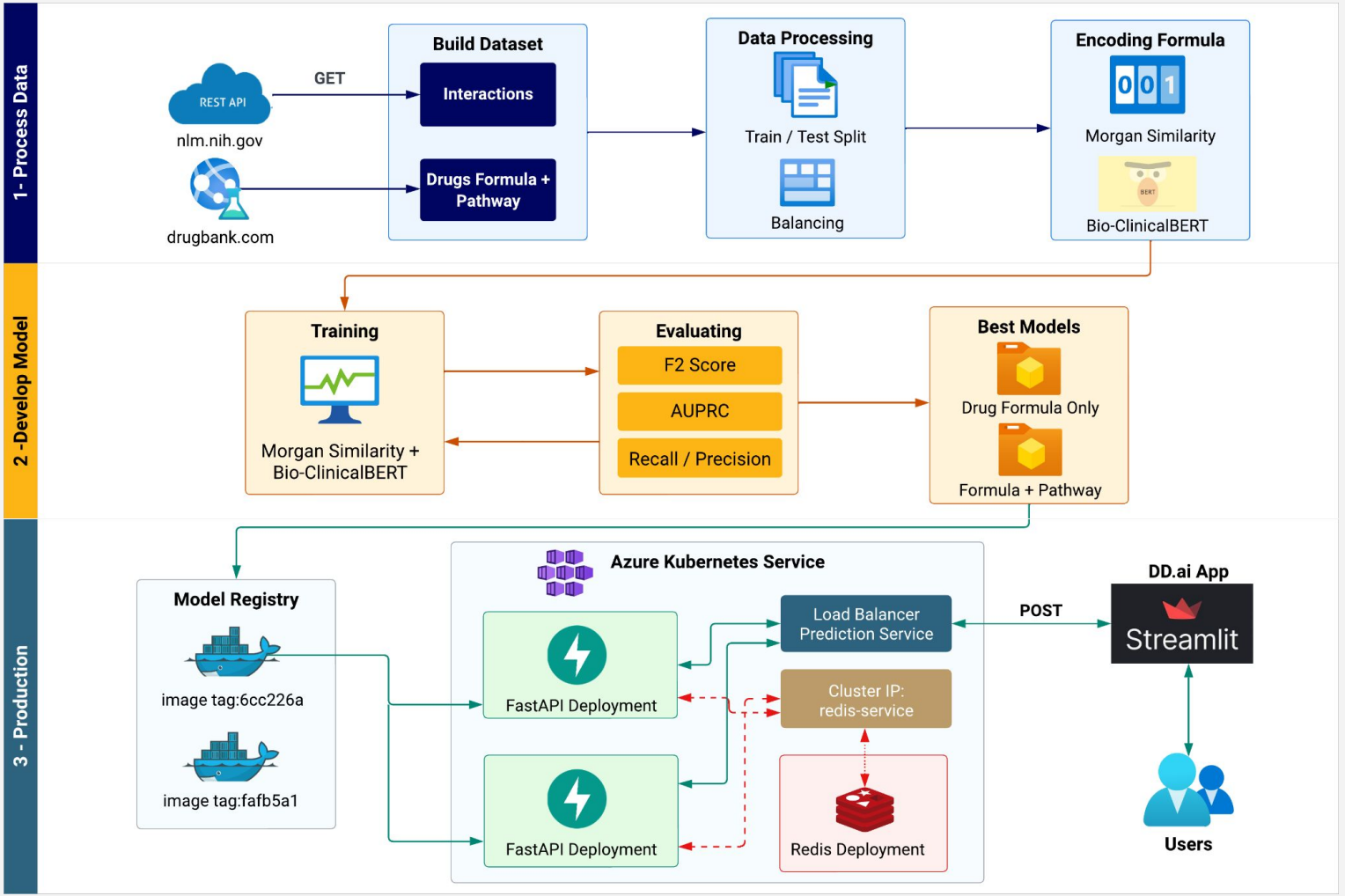
Dr. Tuan La, PharmD
CEO, Global Express Pharmacy

Good start for interaction checker. It would bring greater values to the users by serving a high accurate AI model which focus on a certain drug family like antibiotics or cancer treatment drugs.

03

Approach




1- Process Data
REST API
nlm.nih.gov
GET
drugbank.com
Build Dataset
Interactions
Drugs Formula + Pathway
Data Processing
Train / Test Split
Balancing
Encoding Formula
Morgan Similarity
Bio-ClinicalBERT
2 -Develop Model
Training
Morgan Similarity + Bio-ClinicalBERT
Evaluating
F2 Score
AUPRC
Recall / Precision
Best Models
Drug Formula Only
Formula + Pathway
3 - Production
Model Registry
image tag:6cc226a
image tag:fafb5a1
Azure Kubernetes Service
FastAPI Deployment
FastAPI Deployment
Load Balancer Prediction Service
Cluster IP: redis-service
Redis Deployment
POST
DD.ai App
Streamlit
Users

Data Sources

DrugBank (Features):

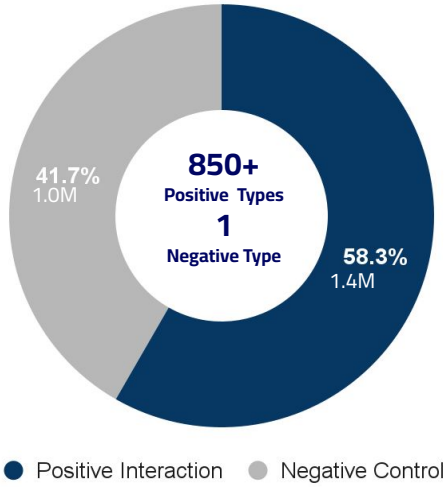
- Drug SMILES
- Action Pathway

Median Word Count / Token Length

2,979 Drugs

NIH (Outcome):

- Drug Interactions



Key Considerations

850+
Interaction Types

Need to reduce scope

SMILES
Chemical structures

Non-traditional text encodings

512
Tokens

Limitation for NLP BERT encodings

Drugbank
data

Limited to pairs of drugs

Scope and Encodings

Scope

850+ Class Prediction Unfeasible

Binary Classification

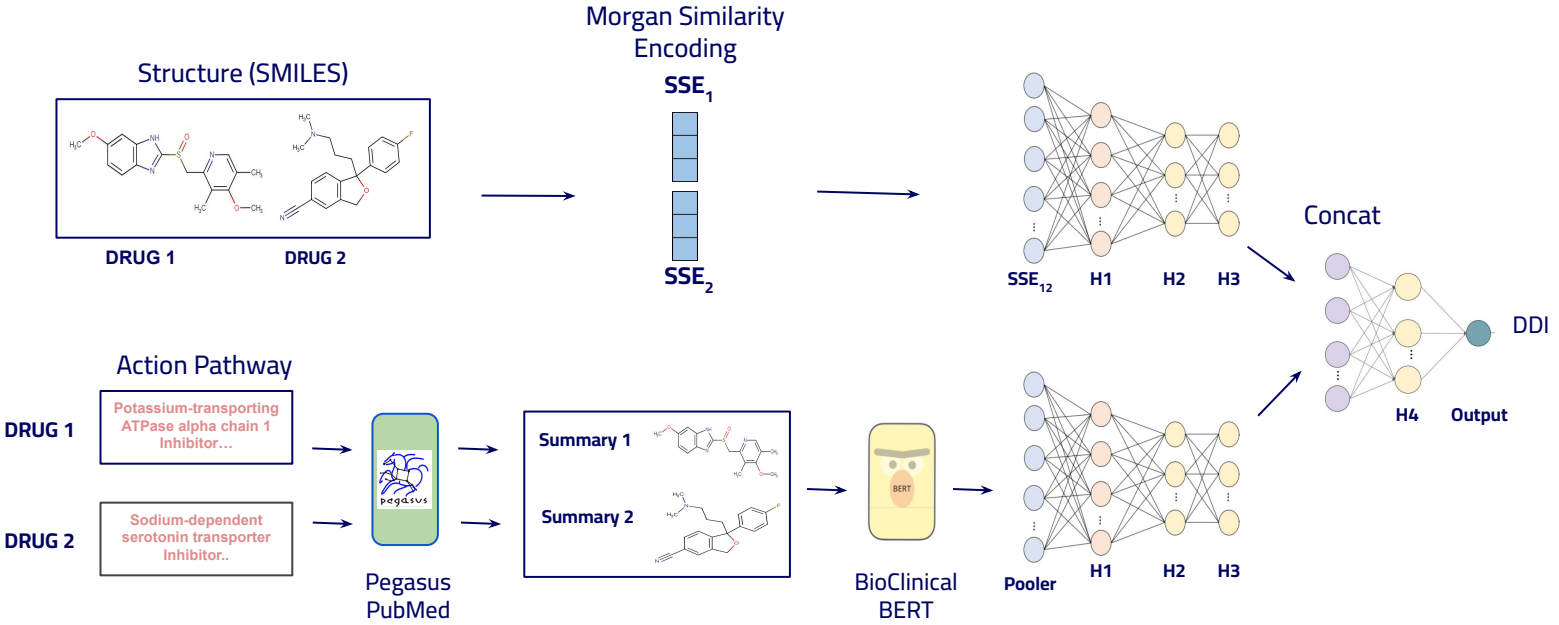
Text-based features

Morgan Similarity
Encode SMILES chemical structures into vectors

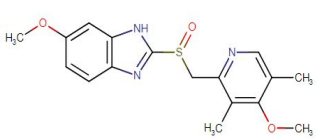
Google Pegasus PubMed
Summarize large text blocks of drug information

BioClinical BERT
Encode above features for enhanced differentiability.

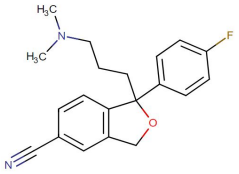
Architecture



DrugA = Omeprazole



DrugB = Citalopram



SMILES (Simplified Molecular-Input Line-Entry System)

`COC1=CC2=C(C=C1)N=C(N2)S(=O)CC1=NC=C(C)C(OC)=C1C`

`CN(C)CCCC1(OCC2=C1C=CC(=C2)C#N)C1=CC=C(F)C=C1`

Morgan Fingerprints Encoding (1024 bits)

`0000010000000010000100100000000010000100010000100...`

`0000000001000000010000001000000000010000100010000100...`

Structure Similarity (SS)

Drug - Basis	Cosine Similarity
SS_{AC1}	0.161

Structure Similarity Encoding (SSE) - 300 centroids from KMeans

`[0.161, 0.226, 0.204, 0.264, ... ]`

$SSE_A = [SS_{AC1}, SS_{AC2}, SS_{AC3}, ...]$

`[0.032, 0.196, 0.206, 0.160,...]`

$SSE_B = [SS_{BC1}, SS_{BC2}, SS_{BC3}, ...]$

Morgan Similarity

Key Takeaways:

- Encode SMILES with Morgan Similarity
- Create 300 clusters of encodings
- Take cosine similarity of new SMILES with 300 clusters to reduce dimensionality

Text Encodings

Google Pegasus PubMed

Summarize large text blocks of drug information

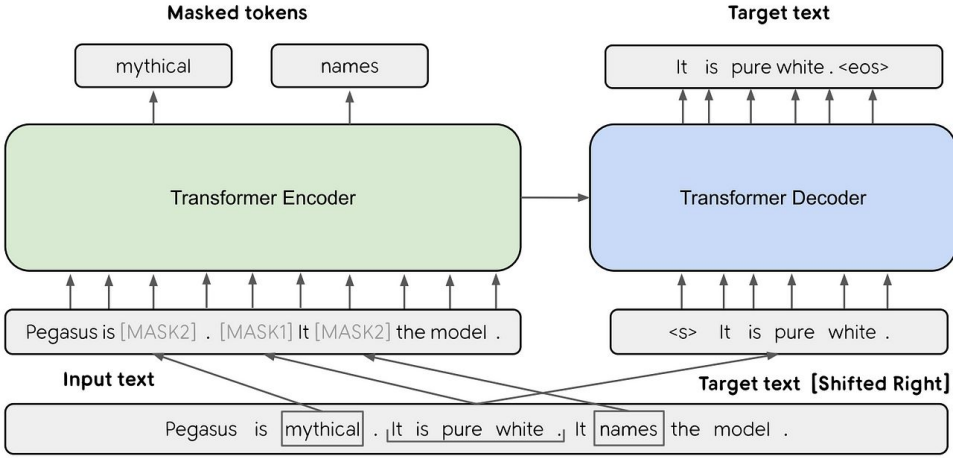


Image Source: https://www.weak-learner.com/blog/2020/06/13/pegasus/

BioClinical BERT

Encode features using BERT model pre-trained on biomedical texts

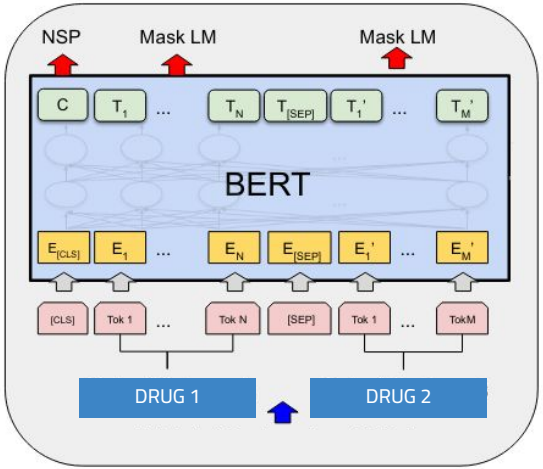
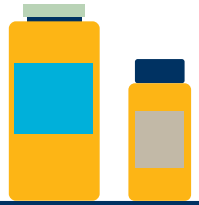


Image Source: https://arxiv.org/pdf/1810.04805.pdf

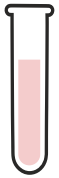
04

Results



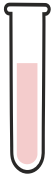
Model Registry

Baselines



Logistic Regression

F2: 70% AUPRC: 72%



XGBoost

F2: 70% AUPRC: 74%

Improvement

Morgan Similarity Neural Net

F2: 73% AUPRC: 83%

BioClinicalBERT SMILES Only

F2: 81% AUPRC: 86%

Final

Morgan Similarity + BioClinicalBERT SMILES Only

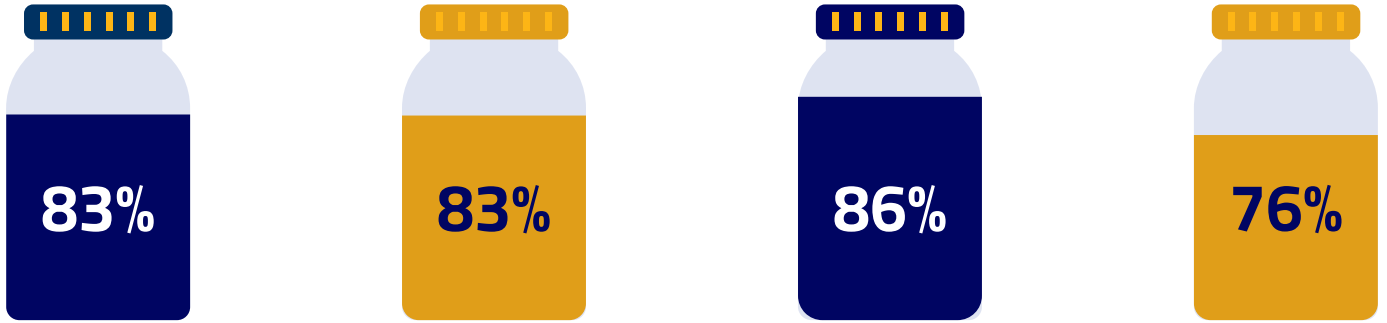
F2: 83% AUPRC: 81%

Morgan Similarity + BioClinicalBERT SMILES & Pathway

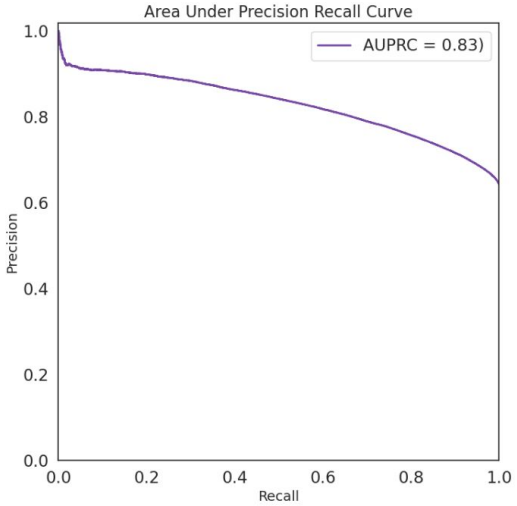
F2: 83% AUPRC: 83%

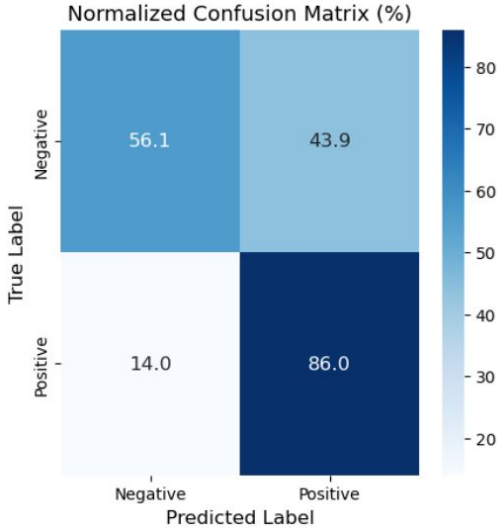
Evaluation

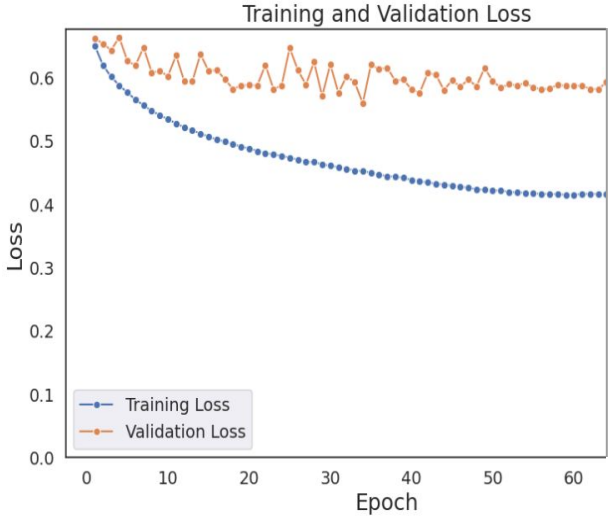
Morgan Similarity + BioClinicalBERT on SMILES & Pathway



Evaluation







Further hypertuning in progress!

Future Developments

Product Focus

Improve model predictions on specific drugs groups (i.e. antibiotics)

Modeling

Graph base GNN
Additional features: target proteins

Encoding

Train BioBERT with drugs target & action pathway texts

Latency

Inference with GPU instances

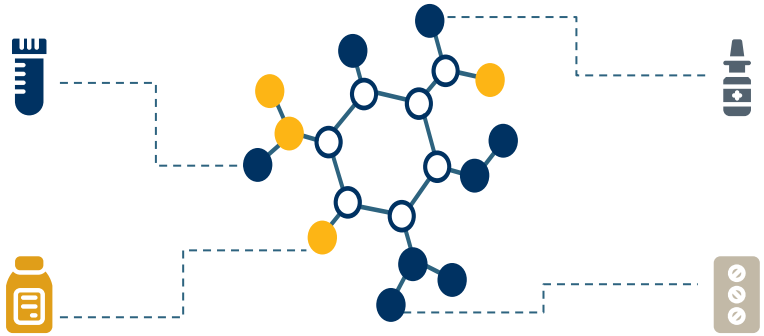
Accomplishments

New Drug Development

Inform potential DDI with drug formula

Scalability

Autoscaling with AKS



NLP Model Serving

Morgan Similarity + BioClinical BERT

Interpretable Results

Graphs and downloadable CSV

Mission

Reimagine drug development by transforming drug-drug interaction detection

Q&A

Thank you!

dd-ai-predict.streamlit.app

Acknowledgements

Professors:

- Joyce Shen
- Cornelia Ilin
- James Winegar

SME, Advisors:

- Roop Raich
- Tim Roth
- Dr. Han Dang, PharmD
- Mai Do, RPh. , MBA
- Dr. Tuan La, PharmD

References

1. Preventable ADE
2. Percent of ADE
3. Adults using multiple medications
4. Drug Discovery Market - Global Industry Analysis, Size, Share, Growth, Trends, Regional Outlook, and Forecast 2023 – 2032
5. Drug Approvals
6. BERTChem-DDI : Improved Drug-Drug Interaction Prediction from text using Chemical Structure Information.
7. Drugbank: a knowledgebase for drugs, drug actions and drug targets
8. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules
9. DrugBank database. Version 5.1.10. 2023
10. Google/Pegasus-Pubmed
11. Publicly Available Clinical BERT Embeddings

Thanks

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and images by **Freepik**