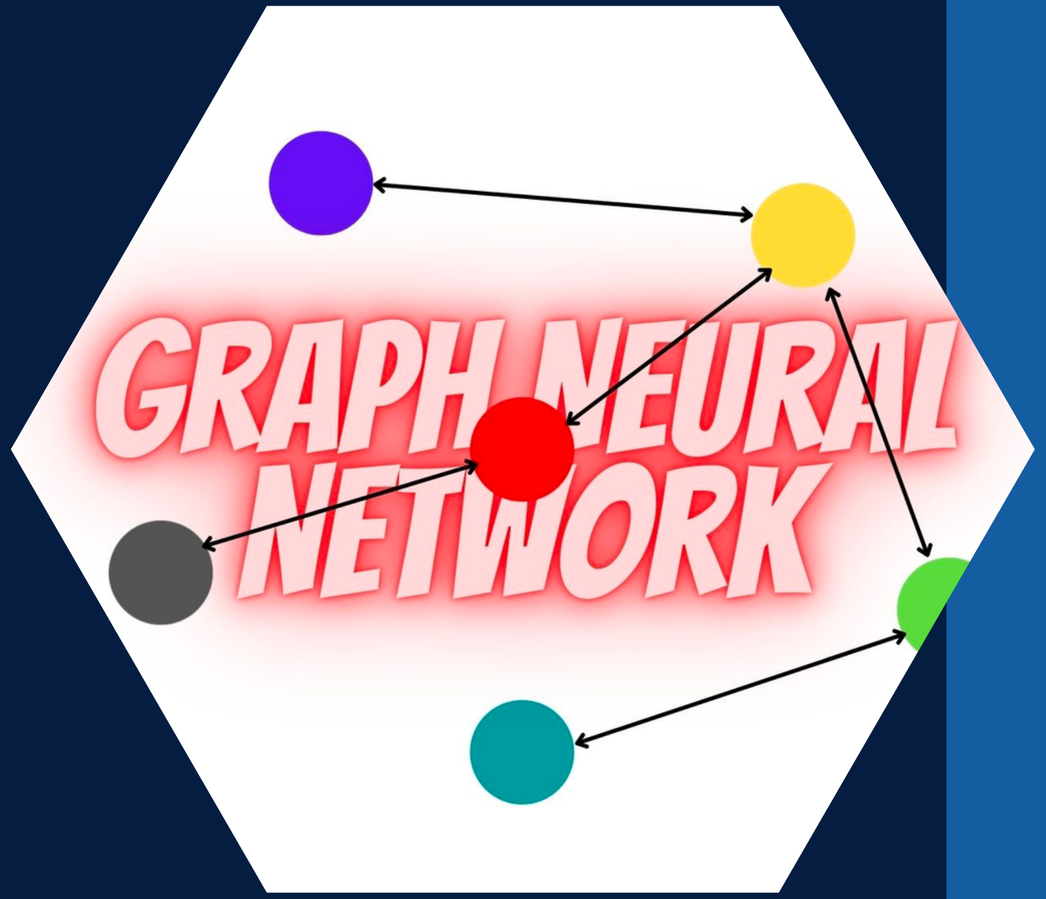
Inspirit AI

DRUG DISCOVERY WITH GNN



Salia Nahshal

Graphs and Neural Networks

- **Deep learning models** like CNN, RNN, and autoencoders have profoundly helped in pattern identification and data mining
- The advanced applications and features of AI and ML enhanced the demand for concepts like GNNs
- Standard methods such as CNNs failed to analyze and represent graphs

What is GNN?

Graph Neural Networks (GNNs) is a deep learning model that infers from a graph data structure where each node learns an embedding containing information about its neighborhood.

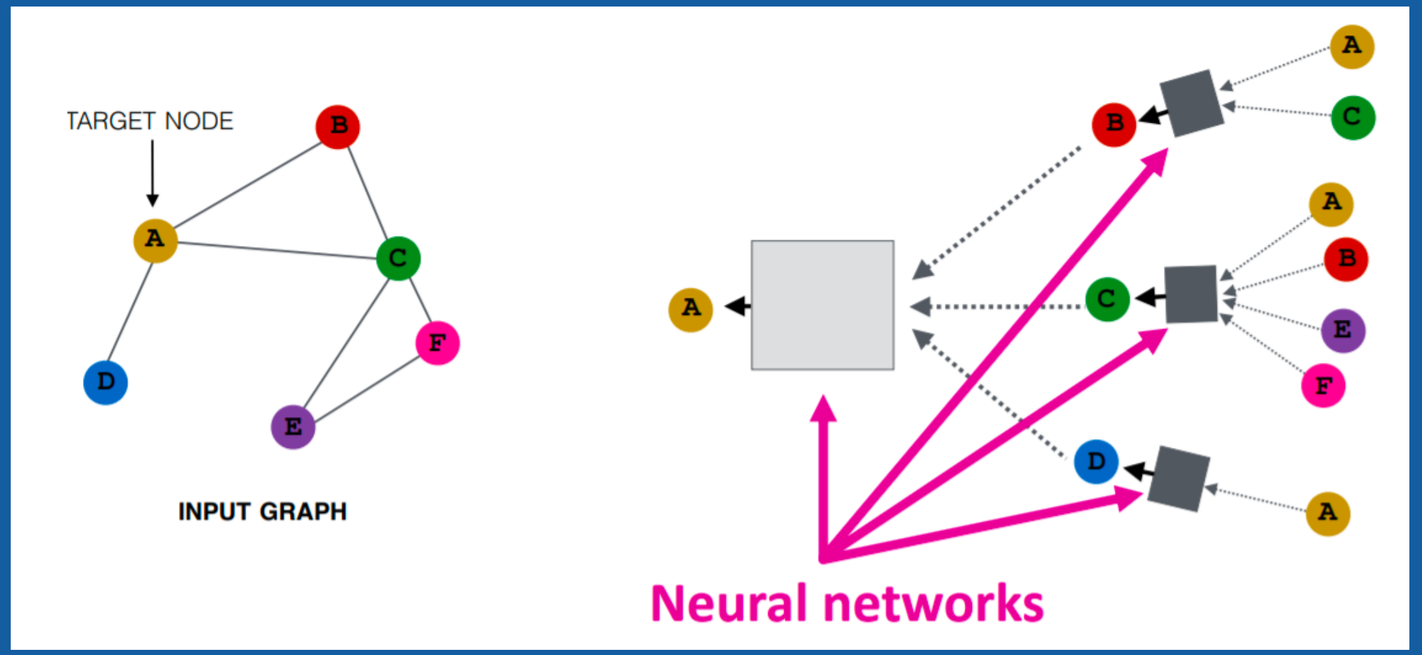


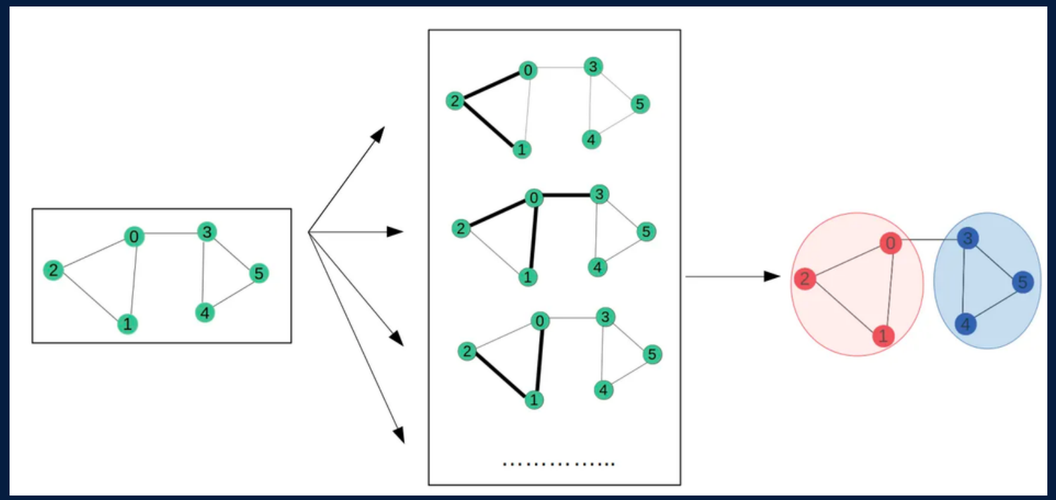
Image by [Purvanshi Mehta]

Types of GNNs

Two types of GNNs are mostly dominant: Graph Convolutional Networks (GCNs) and Graph Auto-Encoder Networks (GAEs).

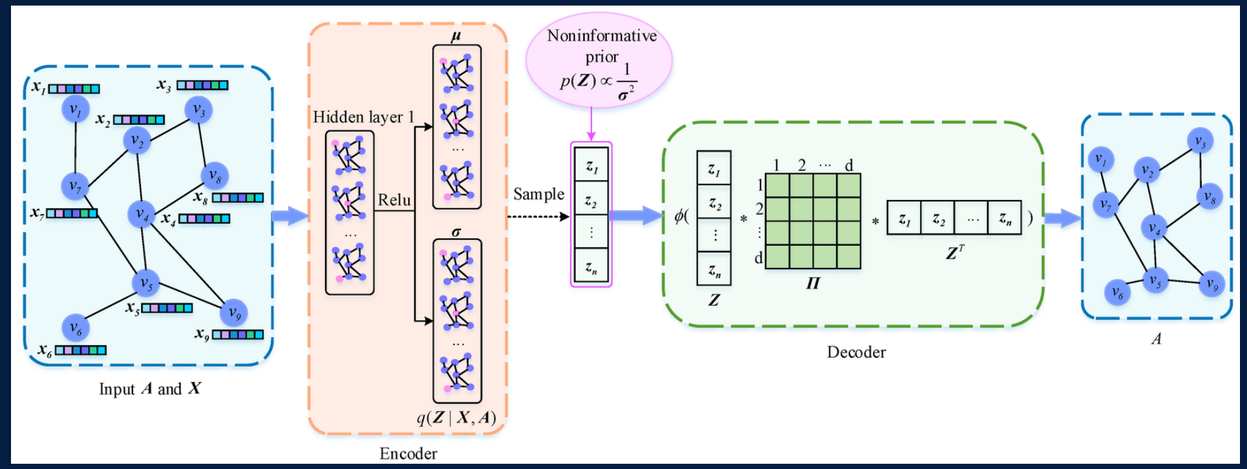
01 GRAPH CONVOLUTIONAL

- GCNs learn features by applying a spatially moving filter on the node of a graph and its neighboring nodes



02 GRAPH AUTO-ENCODER

- GAEs learn graph representation using encoders and attempt to rebuild the input graphs utilizing decoders.



Applications of GNNs

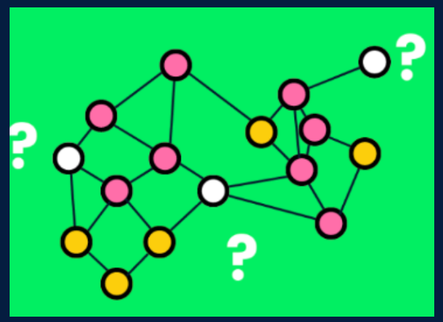
Graph Classification

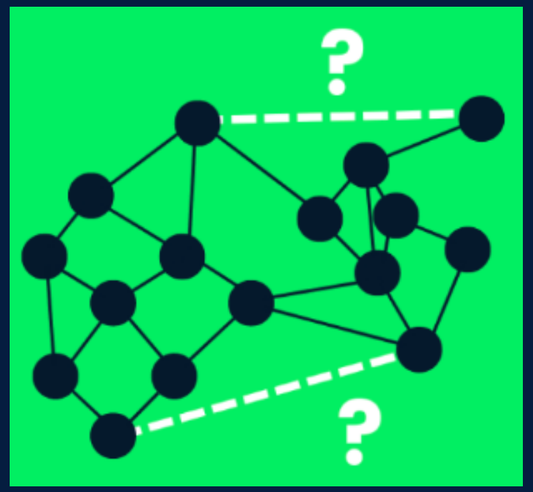
Node Classification

Link Prediction

Graph Clustering

Graph Visualization





GNNs

Advantages

- Captures more information
- Allowed us to work with both structured and unstructured data
  - Ex: complex graph-structured data
- Able to seize non-linear interactions among nodes
- Are inductive

Disadvantages

- Computationally expensive
- Slow to compute
  - Even with GPUs
- If a graph is incomplete, GNNs overfit
- Graph structures constantly change, making it difficult to train a model
- Complicated to implement

Drug Discovery with GNNs

How does drug discovery happen nowadays?

- **Lab experiments**
  - Test compounds against main target until you find what you need
  - Can be accurate but also slow and expensive
- **Mathematical simulations**
  - Use quantum physics to predict chemical properties
  - Fast but less accurate
- **AI modeling with neural networks**
  - Use GNNs
  - Can be fast and highly accurate with the correct model

What is CNN?

Convolutional Neural Networks(CNNs) assist machines in visualizing objects and performing tasks like image recognition, object identification, picture classification, and so on.

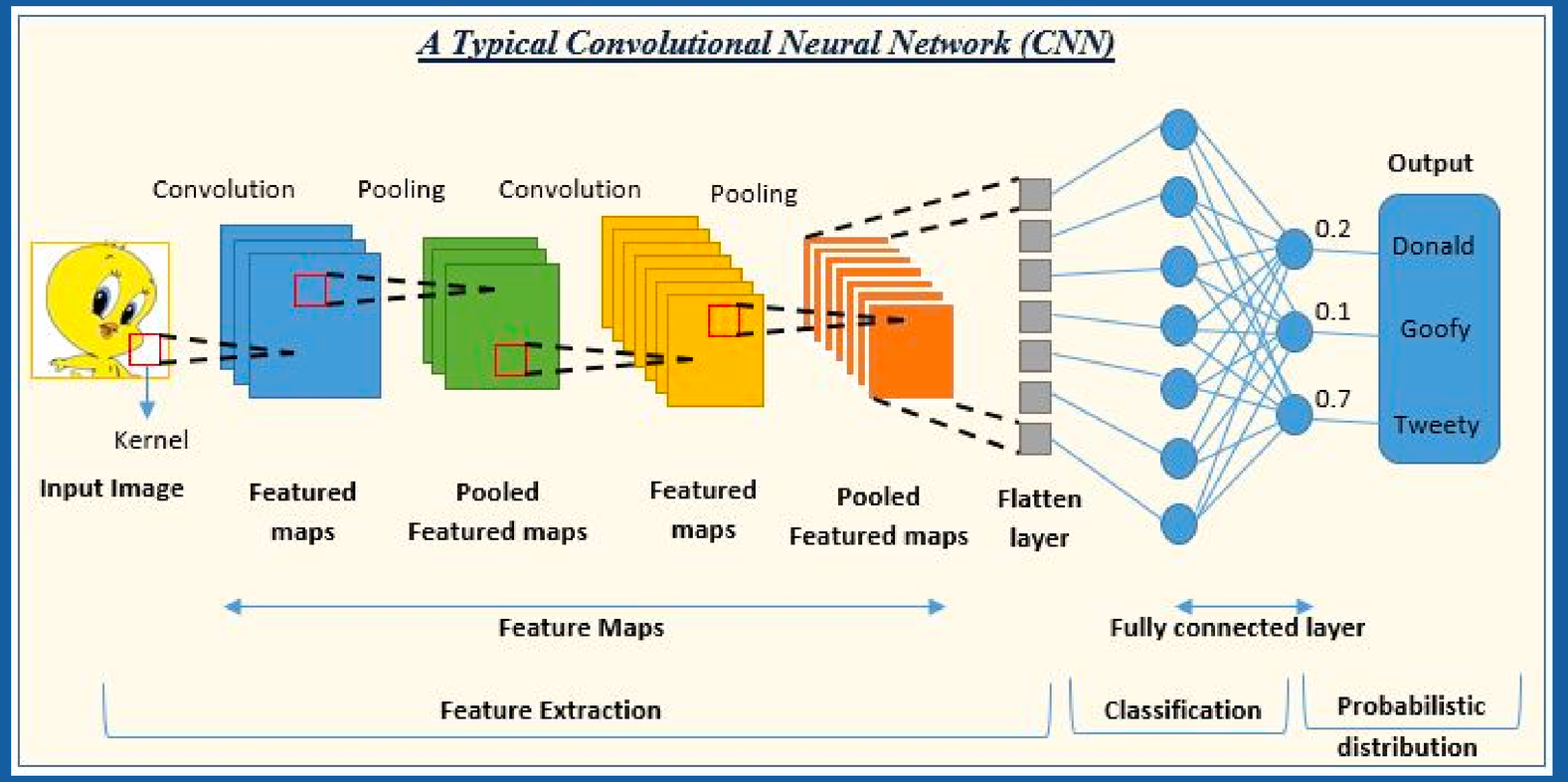
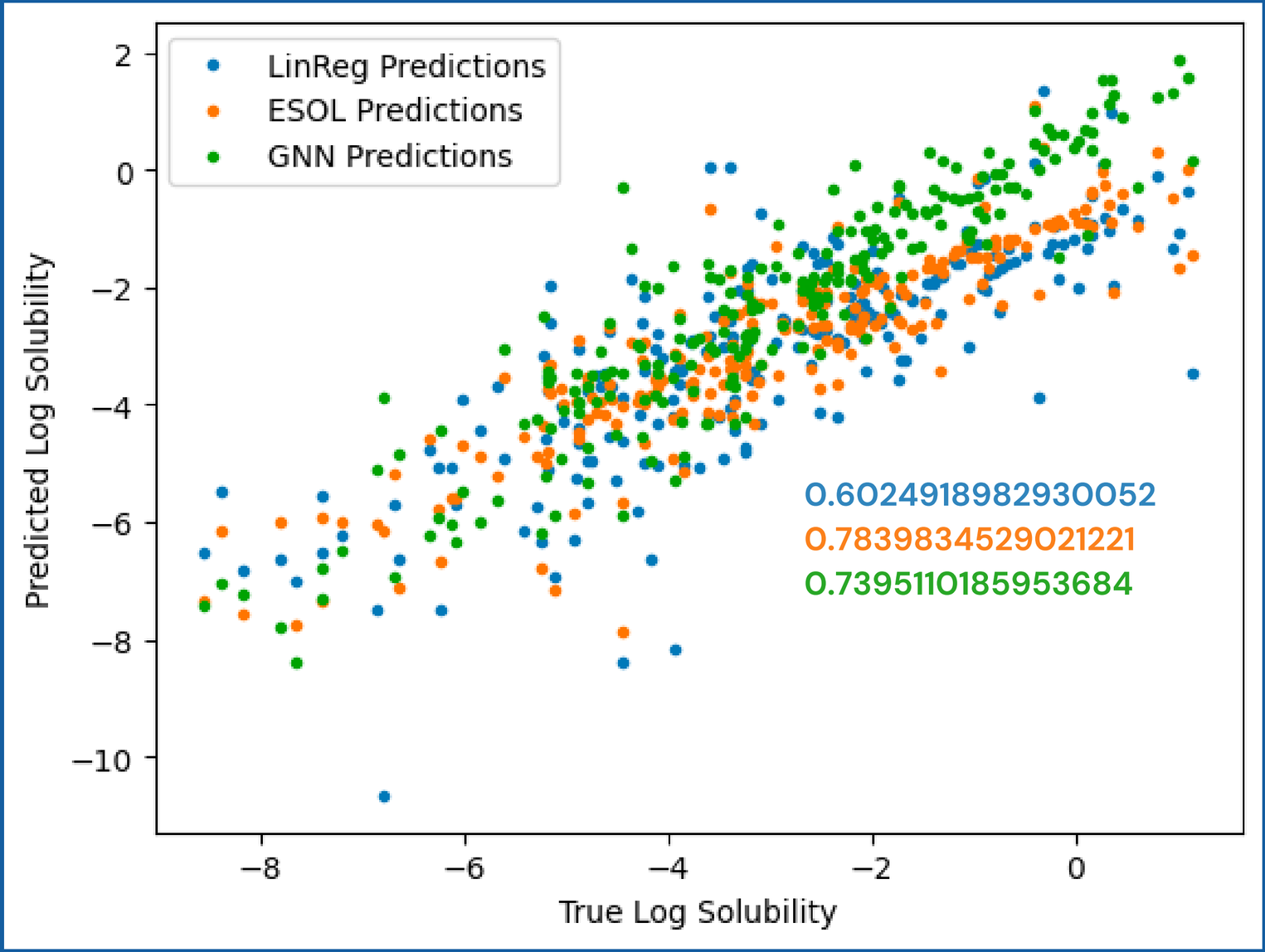


Image by [Analytics Vidhya]

Drug Discovery with GNNs Recap



Toxicity Cont.

- We trained a GNN model for 20 epochs on 12 different takes and a dropout number of 0.2
  - The results were ~0.73

What's Dropout?

In ML, "dropout" is the technique of randomly ignoring specific nodes in a layer during training to reduce overfitting.

Why will dropout help with overfitting?

- It cannot rely on a single input because it may be dropped at random
- Nodes will not learn redundant details of inputs

Toxicity Dropout Rate

Toxicity

```
dropout_values = np.arange(0.1, 1.0, 0.1)
accur_result = []

for d in dropout_values:
    tox_gnn = GraphConvModel(n_tasks=12, mode='classification', dropout=d)
    tox_gnn.fit(tox_train_dataset, nb_epoch=20)
    tox_pred = tox_gnn.predict(tox_test_dataset)[:, :, 1]
    tox_pred_class = (tox_pred > 0.5) # 0 = non-toxic, 1 = toxic
    any_toxic = tox_pred_class.any(axis=1)
    any_toxic_answers = tox_test_dataset.y.any(axis=1)
    accuracy = (any_toxic == any_toxic_answers).mean()
    accur_result.append((d, accuracy))  # Append accuracy to results
```

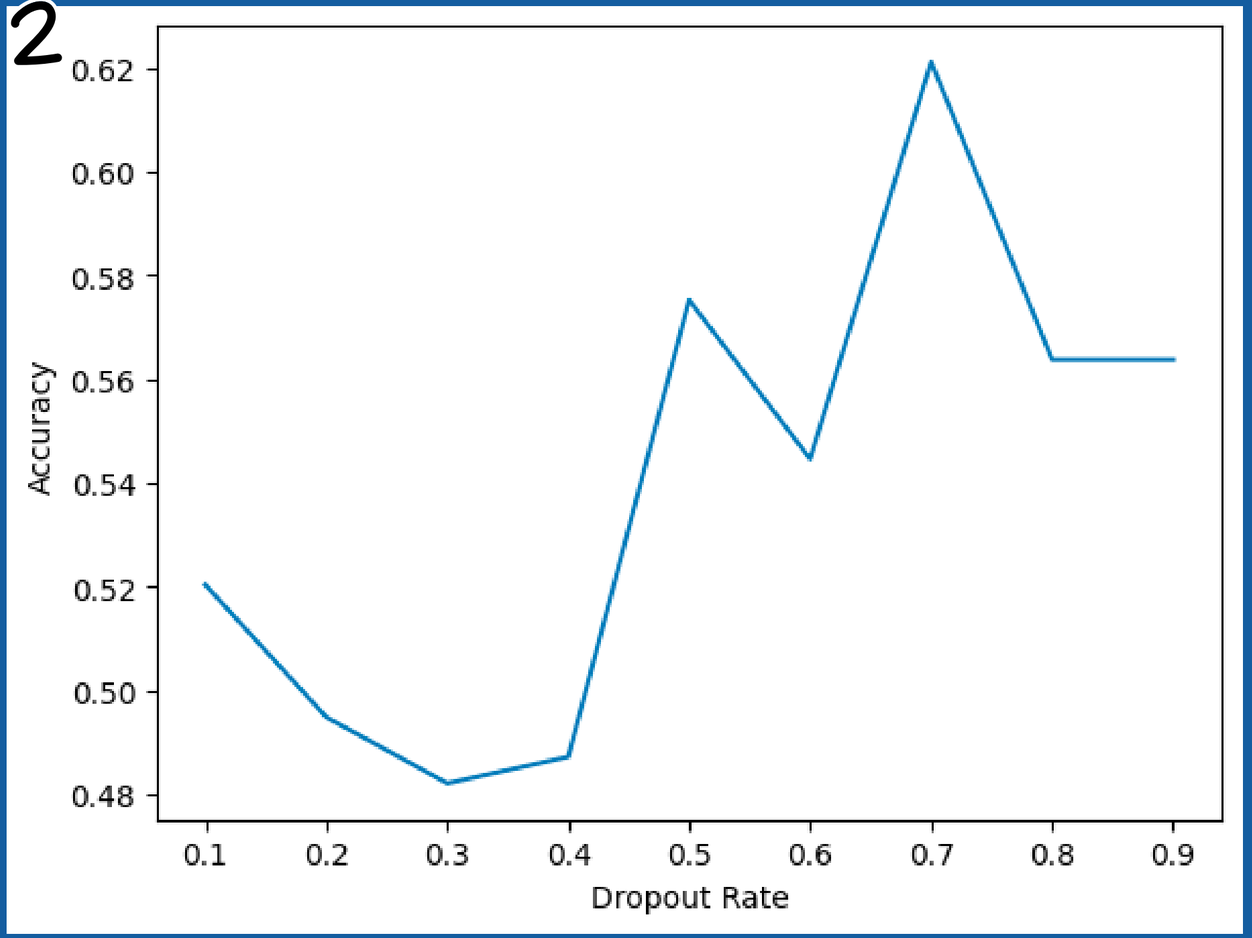
Results

1

Toxicity

```
(0.1, 0.5204081632653061)
(0.2, 0.49489795918367346)
(0.30000000000000004, 0.48214285714285715)
(0.4, 0.4872448979591837)
(0.5, 0.5752551020408163)
(0.6, 0.5446428571428571)
(0.7000000000000001, 0.6211734693877551)
(0.8, 0.5637755102040817)
(0.9, 0.5637755102040817)
```

2



3

Percent of drugs that would pass through **toxicity** and **solubility** screenings: **0.0496**

Thank You!

RESOURCES

- https://www.analyticssteps.com/blogs/introduction-graph-neural-network-gnn
- https://www.analyticssteps.com/blogs/introduction-graph-neural-network-gnn
- https://www.simplilearn.com/what-is-graph-neural-network-article
- https://www.analyticsvidhya.com/blog/2022/08/dropout-regularization-in-deep-learning/#:~:text=Dropout%20is%20a%20regularization%20method,connectedness%20to%20the%20preceding%20layer.
- https://dl.acm.org/doi/10.1145/3487553.3524725
- https://proceedings.neurips.cc/paper/2021/file/b8b2926bd27d4307569ad119b6025f94-Paper.pdf