



Reducing Carbon Emission by Minimizing Steel Waste

ENGM.204-SP24: Data Analytics Project Lab

BY:

KISHORE K, PRANAV D, ROSHNI G, VIBHA V K

Team RelaxedTherm



Roshni Govind

MEM'25

B.Tech Computer
Science (2023)



**Vibha Visvanathan
Kamala**

MEM'25

B.Tech Information
Technology (2023)



**Pranav
Dharmadhikari**

MEM'25

Deloitte USI - Business
Analyst (2022-23)

B.E Chemical
Engineering (2022)



Kishore Kumaran

MSCS'25

Amadeus Labs - SDE 2
(2021-2023)

B.Tech Computer
Science (2021)

Our derived outcomes

Using our Autoencoder-Random Forest model, we were able to predict the strategies that give the maximum utilization of steel sheets to reduce carbon waste with an MAE of ~6%

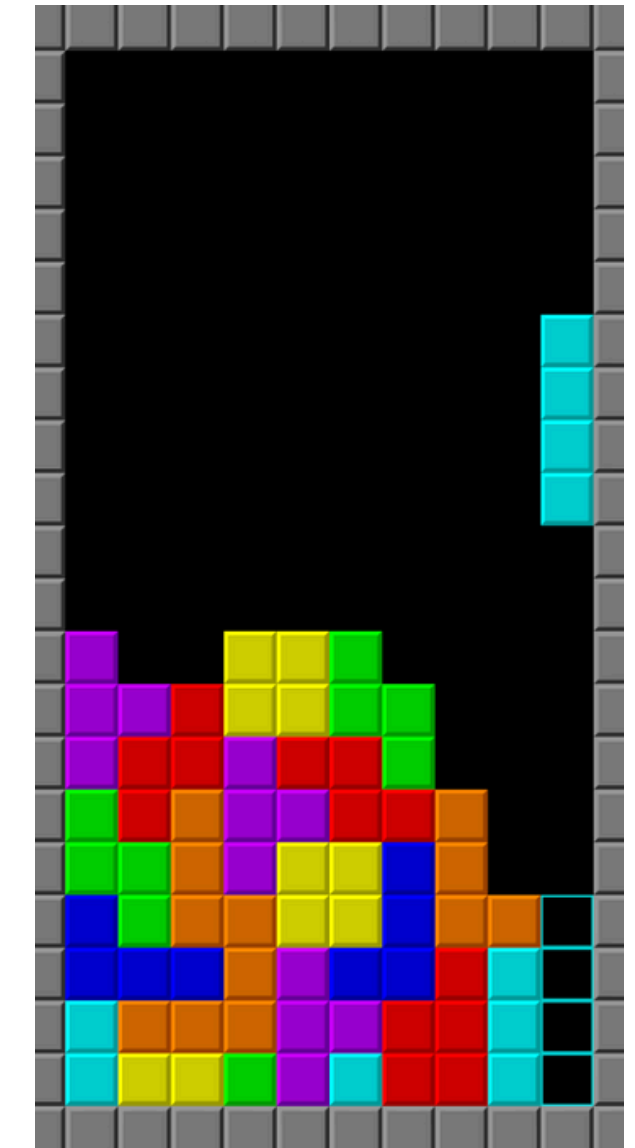
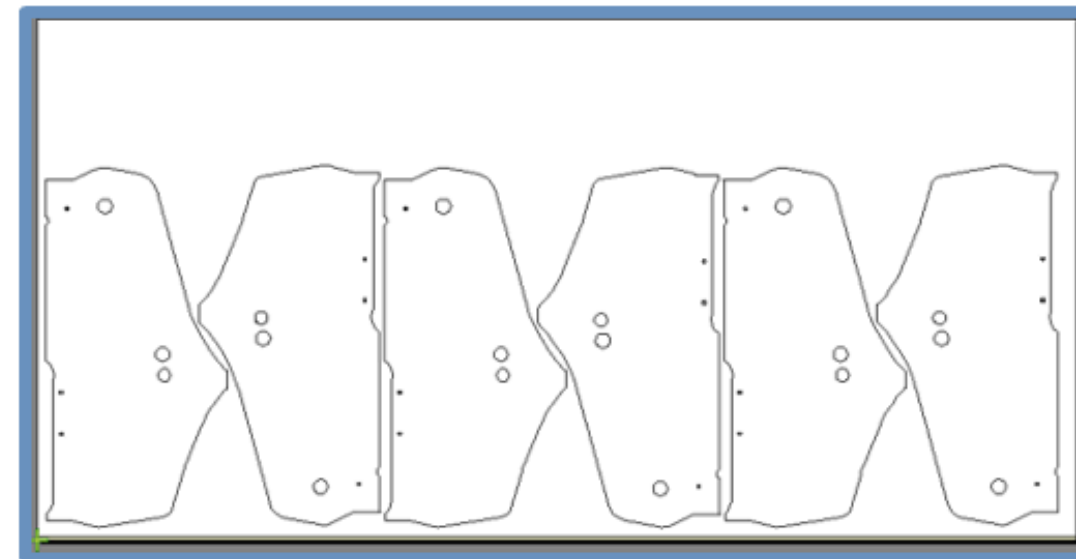
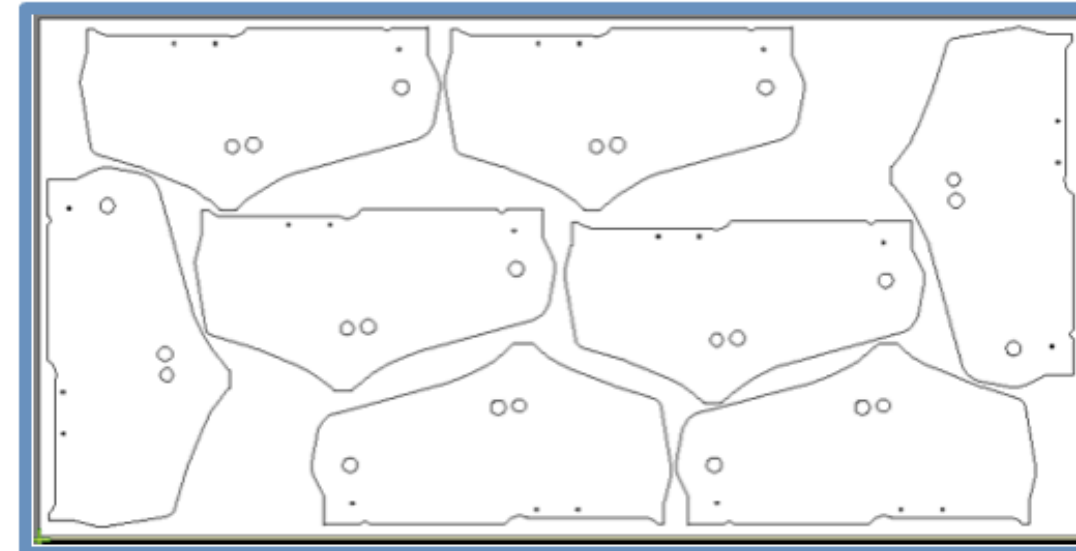
Agenda



- 1 Problem Statement**
- 2 Demo of Result**
- 3 Data Cleaning**
- 4 Data Preprocessing**
- 5 Model Implementation**
- 6 Result and Future Scope**

Hypertherm wants to maximize plate utilization through optimal nesting strategy

- There are 17 strategies available for nesting parts on a steel plate, considering factors such as sheet dimensions and part dimensions.
- Each strategy produces a different nesting pattern, leading to varying levels of sheet utilization.
- The objective is to determine the optimal part placements to maximize the utilization of the sheet.

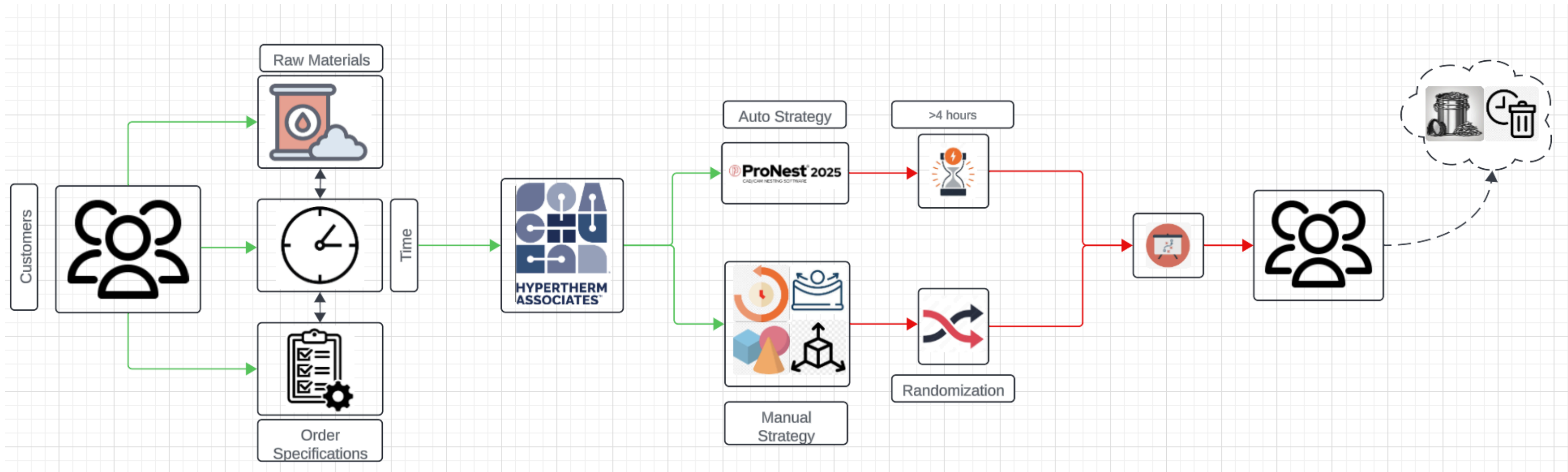


Problem Statement

Hypertherm aims to leverage AI to improve part nesting on steel sheets, exceeding current utilization rates to help customers reduce carbon footprints and cut scrap costs.



We identified two critical pain points in Hypertherm's workflow: time consumed by the software and non-optimal strategy selection, both of which significantly contributed to low raw material utilization.



Demo



```
[35] ✓ 0.0s Python
# Predict utilization for the provided data using all strategies
pred_util_all_strategies = predict_utilization_all_strategies(provided_data)
# for strategy, pred_util in pred_util_all_strategies.items():
#     print(f'Predicted utilization for {strategy}: {pred_util}')

top_strategies = []
for i in range(len(provided_data)):
    # Get predictions for the current row across all strategies
    row_predictions = {strategy: pred_util[i] for strategy, pred_util in pred_util_all_strategies.items()}
    # Sort strategies by predicted utilization
    sorted_strategies = sorted(row_predictions.items(), key=lambda item: item[1], reverse=True)
    # Get the top 3 strategies
    top_3 = sorted_strategies[:3]
    top_strategies.append(top_3)

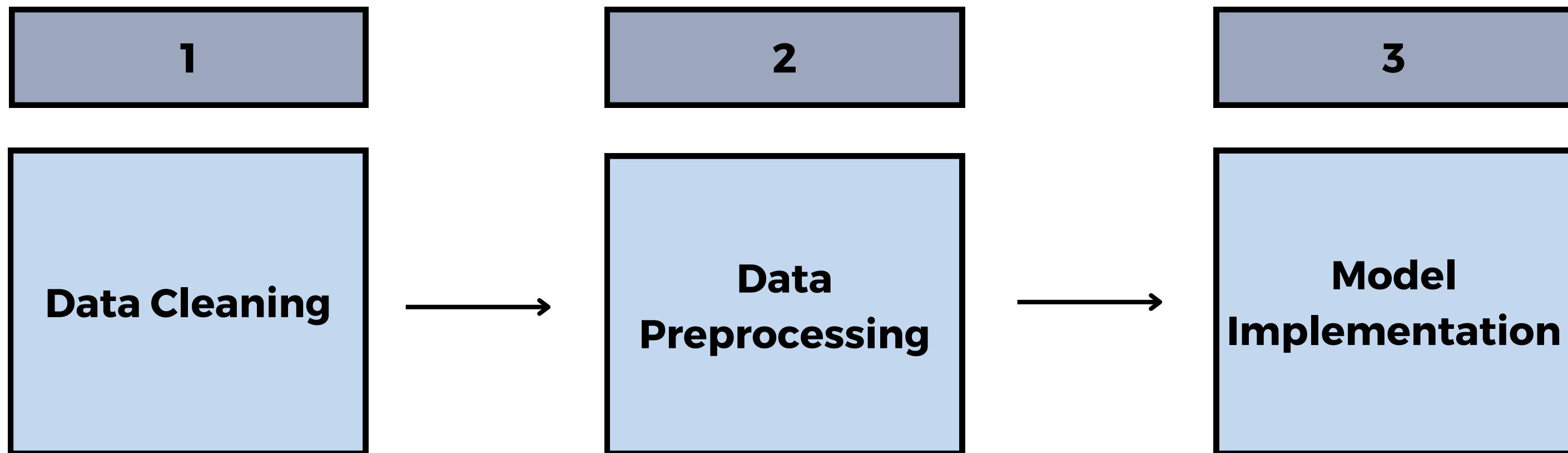
# Print the top 3 strategies for each row
for idx, top_3 in enumerate(top_strategies):
    print(f'Row {idx + 1}:')
    for strategy, util in top_3:
        print(f'  Strategy: {strategy}, Predicted Utilization: {util}')

[36] ↻ 1.5s Python
... 1/1 ██████████ 0s 125ms/step
     1/1 ██████████ 0s 49ms/step
     1/1 ██████████ 0s 61ms/step
     1/1 ██████████ 0s 58ms/step
     1/1 ██████████ 0s 49ms/step
     1/1 ██████████ 0s 65ms/step
     1/1 ██████████ 0s 62ms/step

Spaces: 4  CRLF  Cell 11 of 12
```



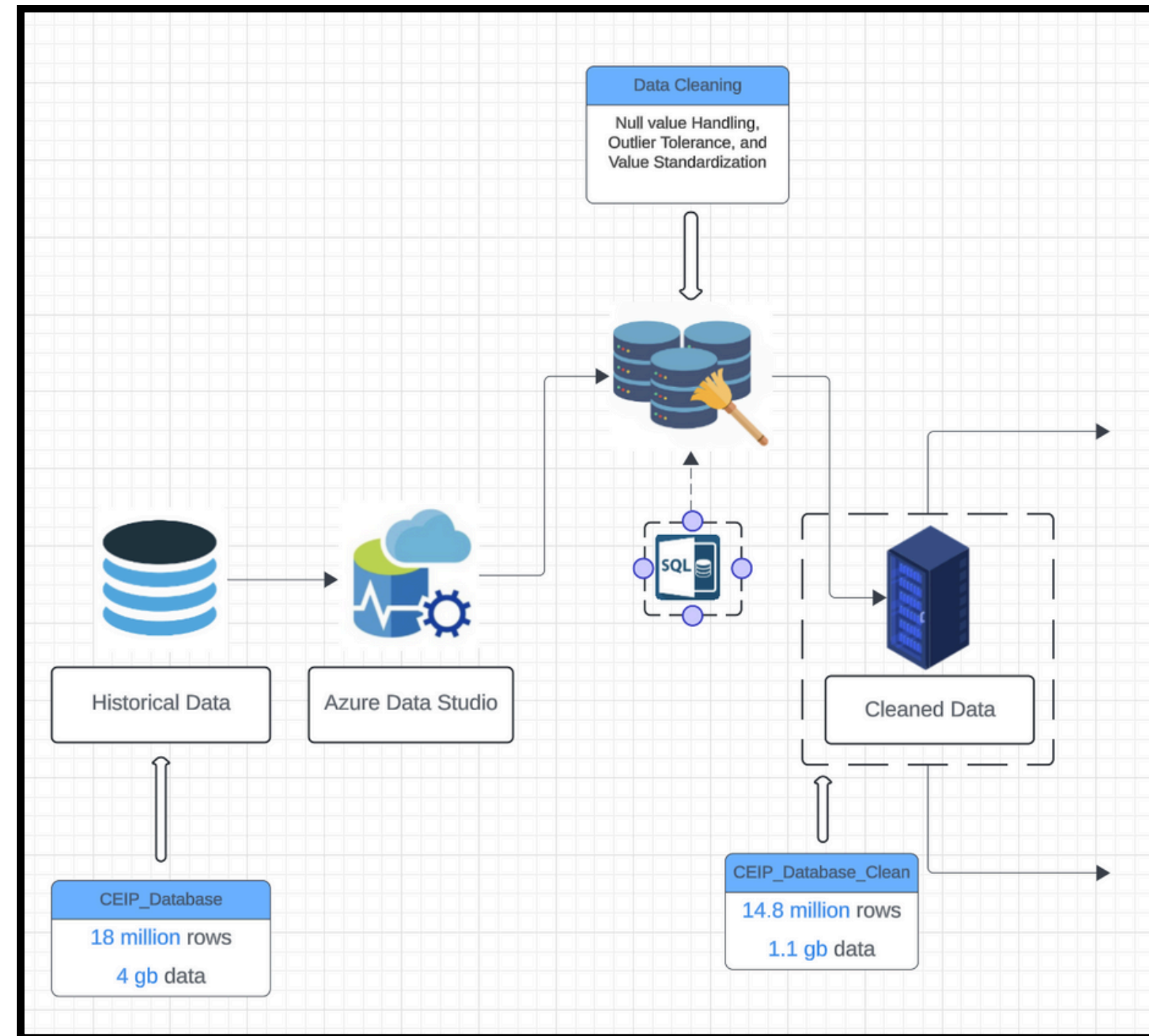
Our solution



Data Cleaning and Preprocessing



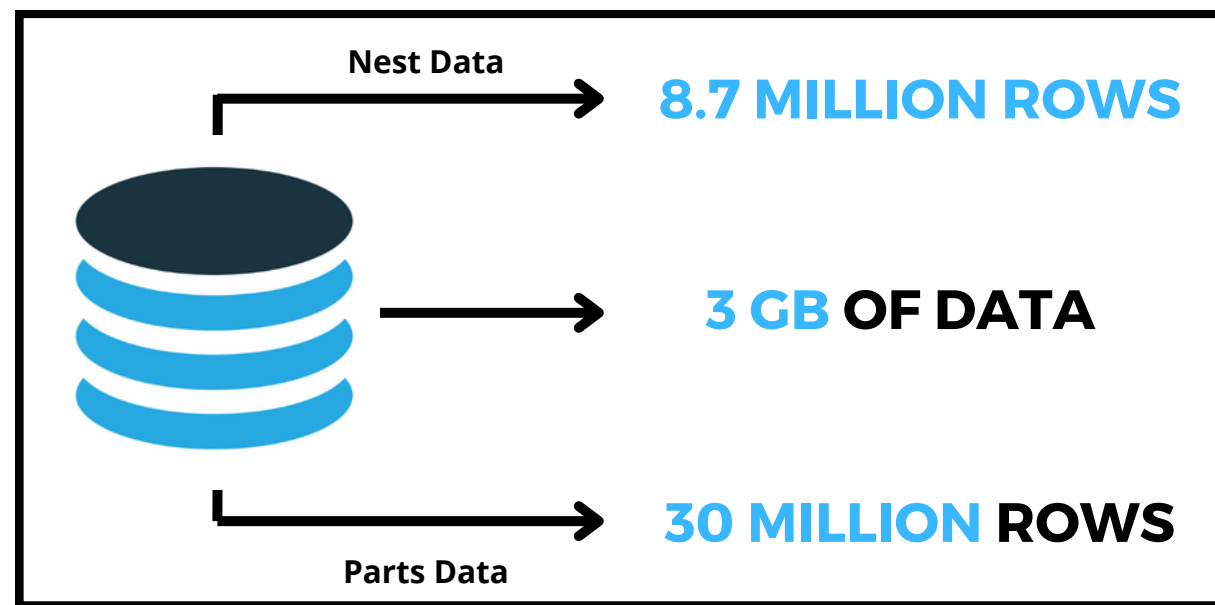
The first step to our Solution was cleaning the Data acquired from Hypertherm and align it to our use case.



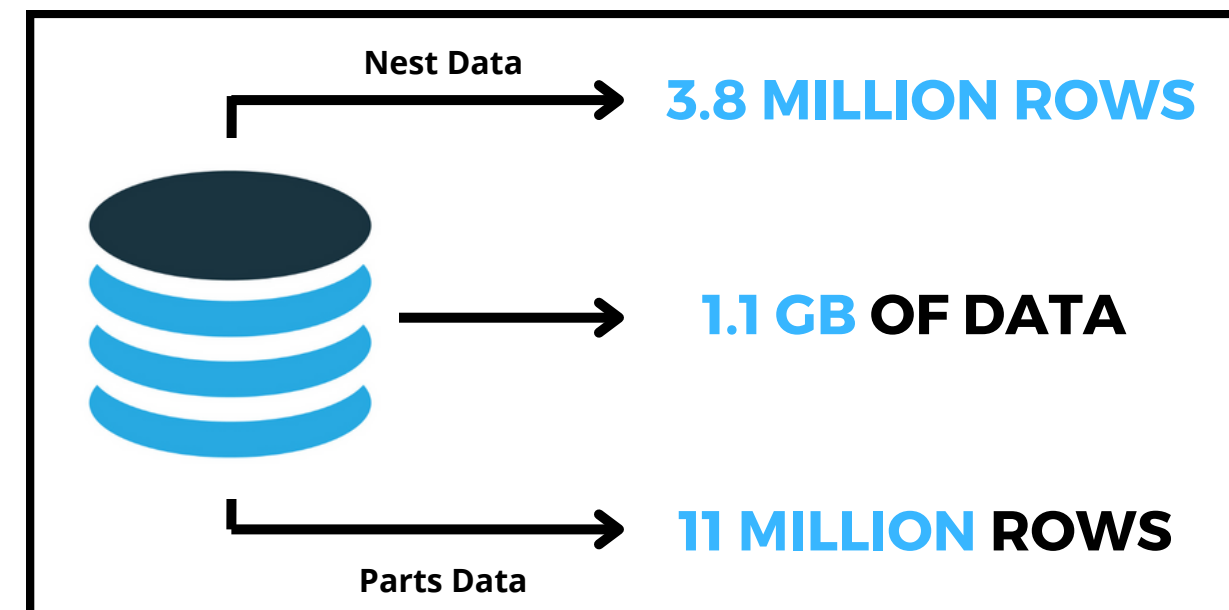
To build our clean dataset we removed Erroneous entries through Null value Handling, Outlier Tolerance, and Value Standardization



Pre-cleaning Metrics



Post-cleaning Metrics



We combined the Parts data and Nest data tables since it was critical for us to aggregate parts dimension and utilisation data to train the model.



Nest Data

Field	Description	Comments
ixJobSummary	Record Index of job containing this nest	
cTimesCut	The number of times the nest will be cut	
fOutput	Has the nest been output?	
cParts	Total number of parts nested	
cSafeZones	Number of safe zones used on the nest	
ixPlateType	Type of plate used	
dNestingTime	Total time spent auto-nesting	
fStrategies	Auto-nesting strategies used	
cMaxTorches	Maximum number of torches on nest	
dMaxTorchSpacing	Maximum torch spacing used on nest	
dLength	Sheet length	
dWidth	Sheet width	
dArea	Sheet area	
ixMaterial	Record index of material used	
dLengthUsed	Length of plate used by nested parts	
dWidthUsed	Width of plate used by nested parts	
dCropUtil	Nested utilization of parts inside of crop (if one exists)	Nested part area / (sheet area – area of remnants saved from nest)
dPartArea	Total area of nested parts	
dTrueArea	Plate area used by nested parts	True area of sheet. For skeletons, this excludes the area of the cutouts.

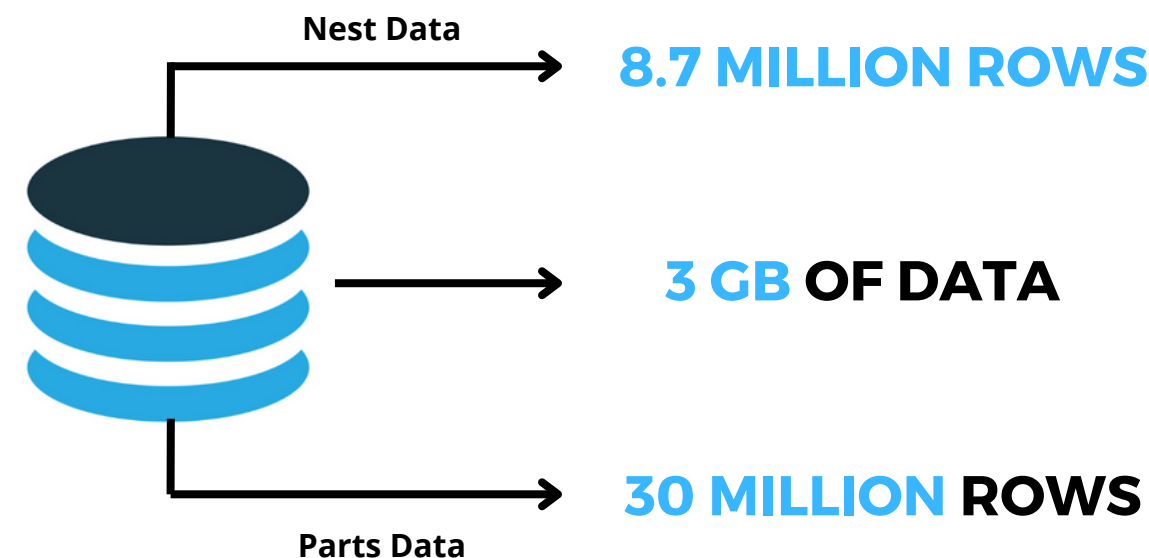
Part Data

Field	Description	Comments
ixPart	Part record index	
ixJobSummary	Record Index of job containing this part	
dLength	Part length	
dWidth	Part width	
dArea	Part true area	Area of exterior profile – area of cutouts
cRequired	Number of parts required	For the entire job
cNested	Number of parts nested	For the entire job
ixMaterial	Record index of material used	
fExtShape	Shape of the exterior profile	About 50 known shapes
dExtArea	Area of the exterior profile	
dExtBoundaryDist	The maximum distance of any point inside the profile to the nearest point on the exterior profile	Uses a distance transform with distance measured at 45 degree increments
dExtContainedDist	The maximum unbroken distance between any two points on the exterior profile	The line between the two points doesn't intersect the profile anywhere else
dLgIntArea	Area of the largest interior profile	
dLgIntBoundaryDist	The maximum distance of any point inside the profile to the nearest point on the interior profile	Uses a distance transform with distance measured at 45 degree increments
dLgIntContainedDist	The maximum unbroken distance between any two points on the interior profile	The line between the two points doesn't intersect the profile anywhere else
dLgExtConArea	Area of the largest concavity	
dLgExtConBoundaryDist	The maximum distance of any point inside the concavity to the nearest point on the concavity	Uses a distance transform with distance measured at 45 degree increments
dLgExtConContainedDist	The maximum unbroken distance between any two points on the concavity contour	The line between the two points doesn't intersect the profile anywhere else

Why ixJobSummary?

Consistency and Data Consolidation: The ixJobSummary metric remained consistent across both tables, facilitating a unified dataset for reliable comparison and analysis.

Model Training Efficiency: It provided a high-level overview and helped streamline the model training process by consolidating data from multiple sources.



We detected the following Anomalies while cleaning the data.

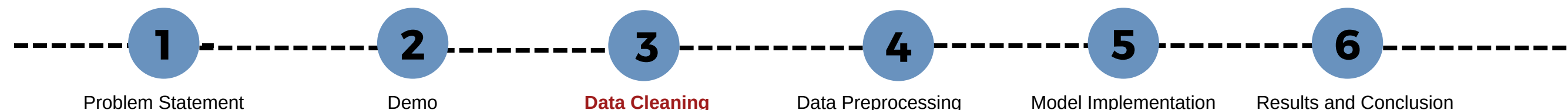
Nest Table:

- Rows with $dCropUtil = 0$
- Rows with $dPartArea = 0$
- Rows with $dTrueArea = 0$
- Rows with $dLengthUsed = 0$ and $dLengthUsed < 0$
- Rows with $dWidthUsed = 0$ and $dWidthUsed < 0$
- Rows with $cParts = 0$
- Rows with $fStrategies$ outside of 17 strategies defined in the problem statement
- In some rows, $dLengthUsed$ and $dWidthUsed$ is more than the $dLength$ and $dWidth$, so we selected a tolerance of 2 and deleting rows where the deviation is more than this, and correcting the values within this tolerance
- $dCropUtil$ calculated based on $dTrueArea$ for over 60% of data and based on $dArea$ for the remaining records. Standardized to calculation based on $dTrueArea$ based on feedback from Hypertherm
- Rows where $dPartArea$ is greater than $dTrueArea$

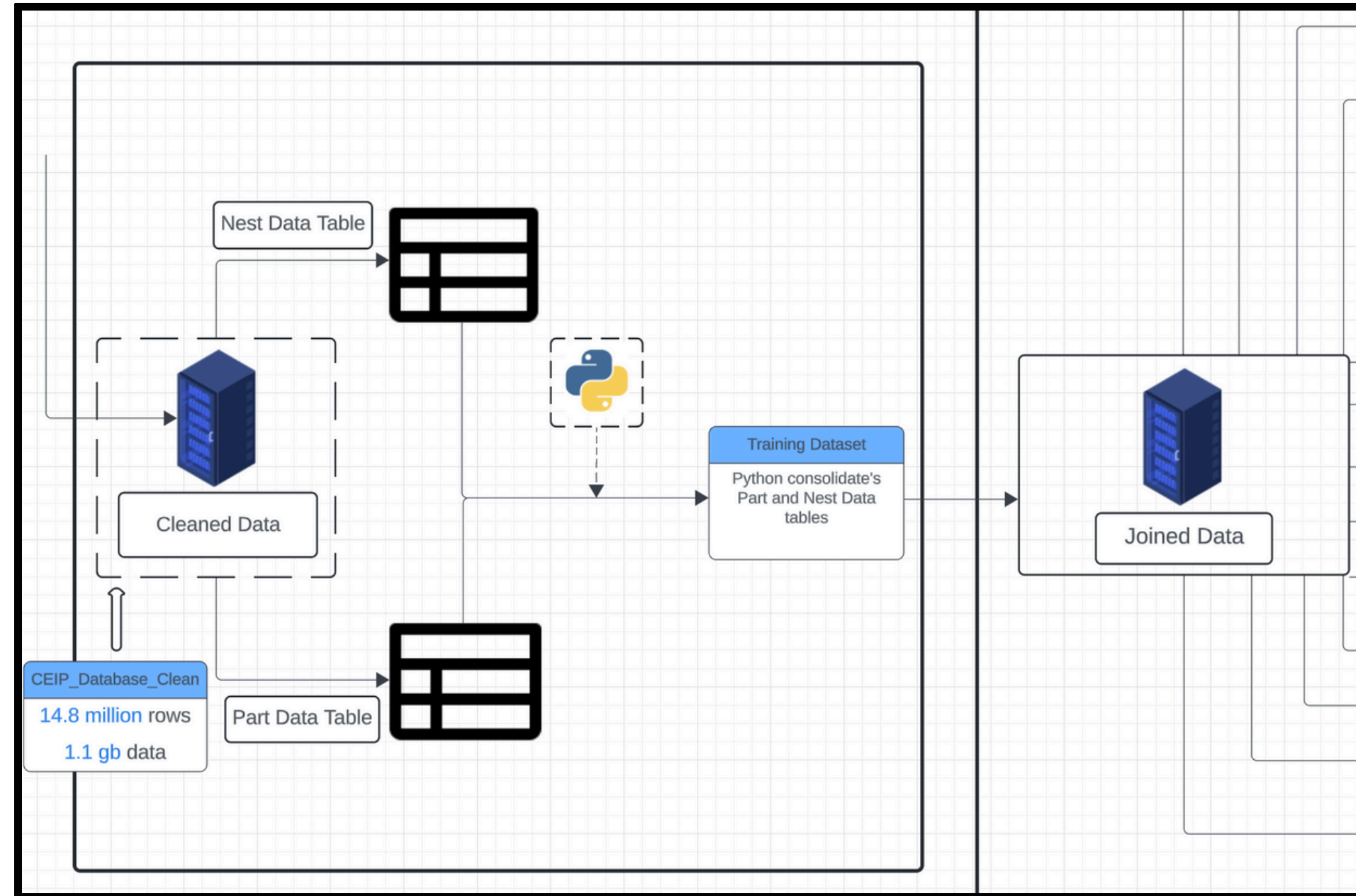
Part Table:

- Rows with $dLength = 0$ and $dLength < 0$
- Rows with $dWidth = 0$ and $dWidth < 0$
- Rows with $dArea = 0$ and $dArea < 0$
- Rows with $cNested = 0$ and $cNested < 0$

Finally, using number of parts and area of each part in a job from the Part table and matching it with the number of nested parts in all the nests within a job and area of nested parts within a job to ensure synergy between Part and Nest table.



The next step involved aggregating the cleaned data to gain a comprehensive overview of the sheet and part specifications, which allowed for more effective model training.



Model Implementation

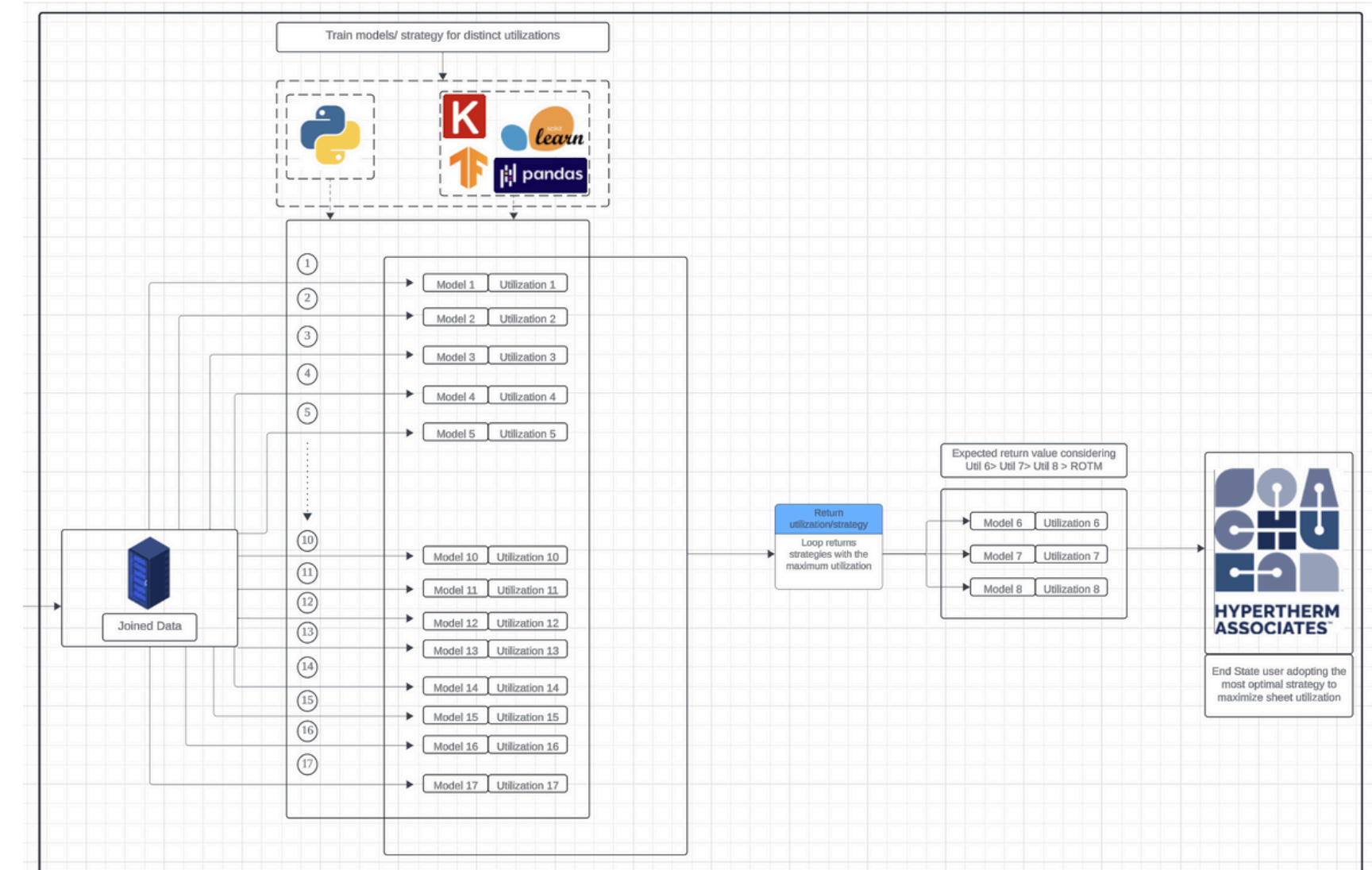
Strategy-specific models improve utilization predictions by learning nuances and preventing cross-strategy interference

Training Individual Models for Each Strategy:

- Each model learns the particular nuances of its assigned strategy
- By capturing unique data patterns and within each strategy, the models can provide less MAE
- Segregating models prevents interference between different data distributions.

Prediction Workflow for New Data

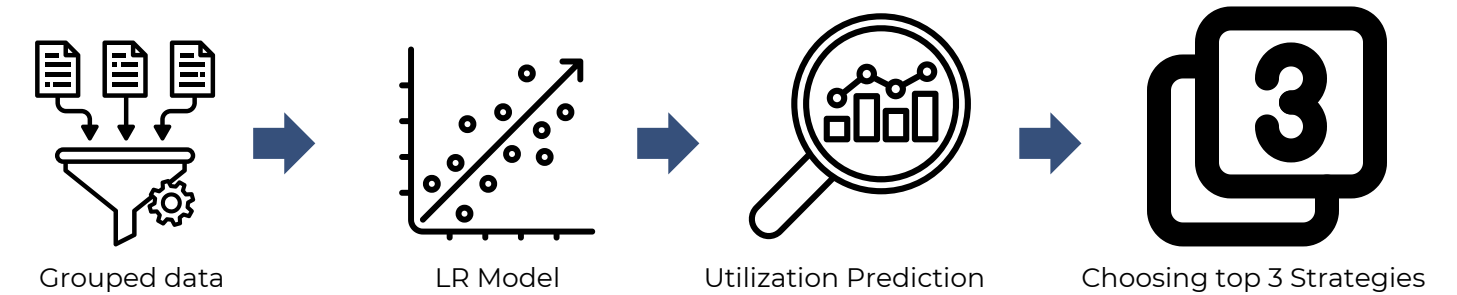
- The standardized data is fed into the corresponding model tailored for each strategy
- Predictions of `CropUtil` are collected from all the strategy-specific models.
- The predictions are then ranked to identify the top 3 strategies with the highest predicted utilization.



Model 1: Linear Regression Model

Why Linear Regression?

Most Commonly used model for predictions, easy to implement, baseline model



How the Model Works:

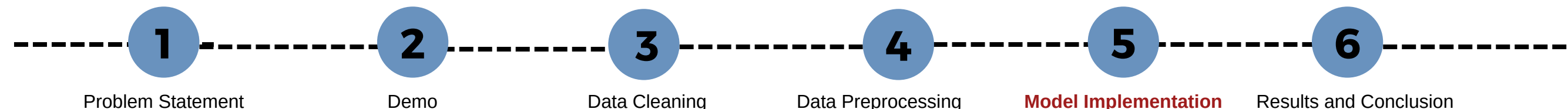
The data is grouped by different strategies, resulting in separate datasets for each strategy. We train individual linear regression models on these strategy-specific datasets to predict utilization tailored to each strategy. Each model is then evaluated with new data, and final utilization predictions are generated accordingly

Performance:

The average MAE for the LR models of all the strategies averages out to 11.36%.

```

Mean Absolute Error for strategy -2147483648: 0.17243847310179375
Mean Absolute Error for strategy 0: 0.17007085061048358
Mean Absolute Error for strategy 1: 0.12087709037506864
Mean Absolute Error for strategy 2: 0.07587781897573186
Mean Absolute Error for strategy 4: 0.10831671082297177
Mean Absolute Error for strategy 8: 0.11130536659789601
Mean Absolute Error for strategy 16: 0.1008931848941399
Mean Absolute Error for strategy 32: 0.11463965895750546
Mean Absolute Error for strategy 64: 0.10674503232528745
Mean Absolute Error for strategy 128: 0.11334468477553149
Mean Absolute Error for strategy 256: 0.09056337047297233
Mean Absolute Error for strategy 512: 0.11708796531329731
Mean Absolute Error for strategy 1024: 0.12435057973973145
Mean Absolute Error for strategy 2048: 0.12670126339086432
Mean Absolute Error for strategy 4096: 0.12620449579268003
Mean Absolute Error for strategy 8192: 0.1336700162174164
Mean Absolute Error for strategy 16384: 0.1184001536864067
  
```



Model 2: Artificial Neural Networks (ANN)

Why ANN?

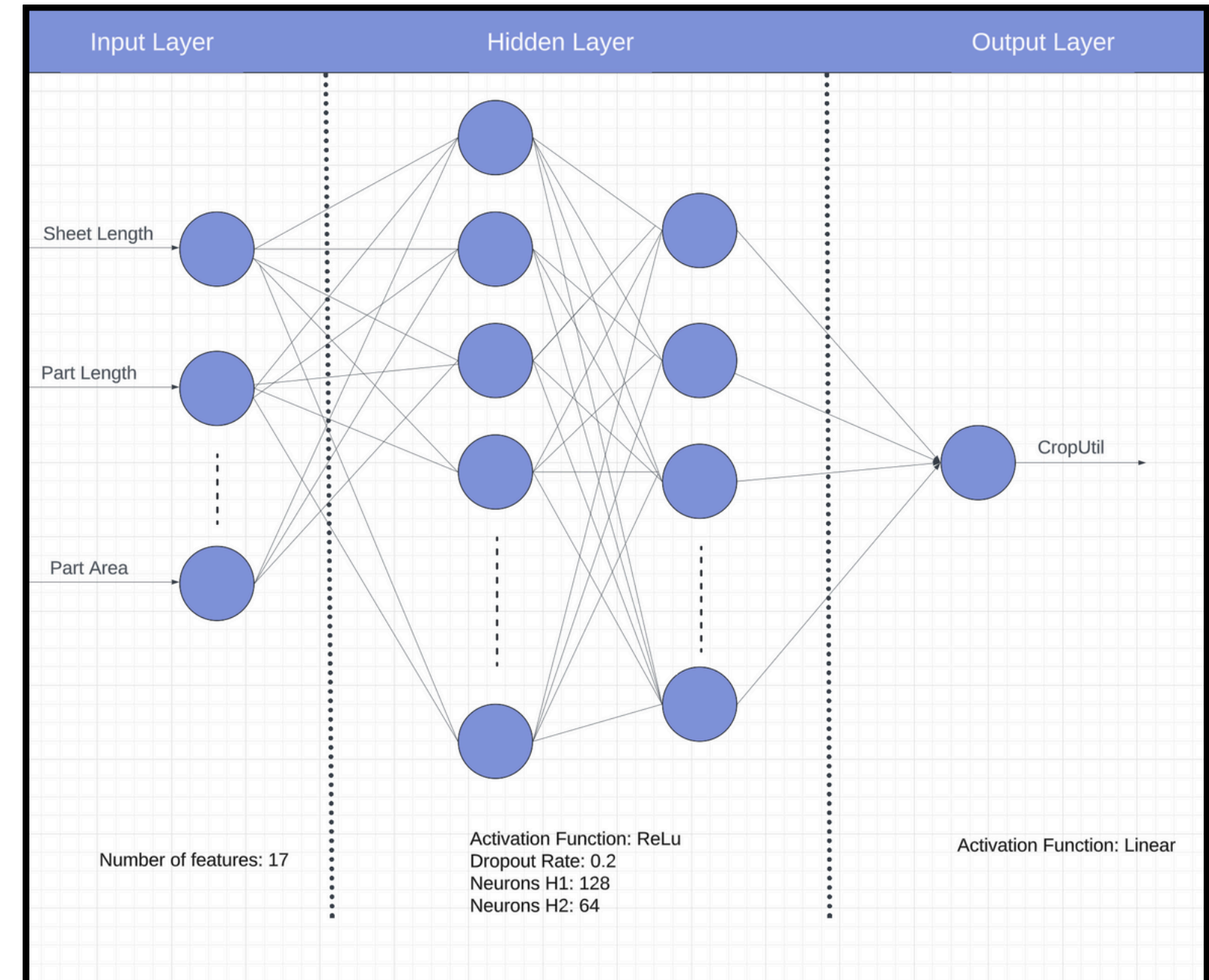
ANNs can automatically learn and extract significant features from the raw input data through the hidden layers, reducing the need for manual feature engineering.

How the Model works:

The combined dataset is divided by strategies for separate modeling. The ANN architecture includes an input layer with neurons equal to the number of features, two hidden layers with 128 and 64 neurons (ReLU activation and 20% dropout), and an output layer with one neuron using a linear activation for regression. Key callbacks like `EarlyStopping` and `ReduceLROnPlateau` prevent overfitting and adjust the learning rate. Model performance is evaluated using Mean Absolute Error (MAE). Finally, the trained model makes `CropUtil` predictions on new, standardized data.

Performance:

The average MAE for the ANN models of all the strategies average out to 8.17%.



Model 3: Autoencoder-Random Forest Model

Why the Ensemble Autoencoder-Random forest?

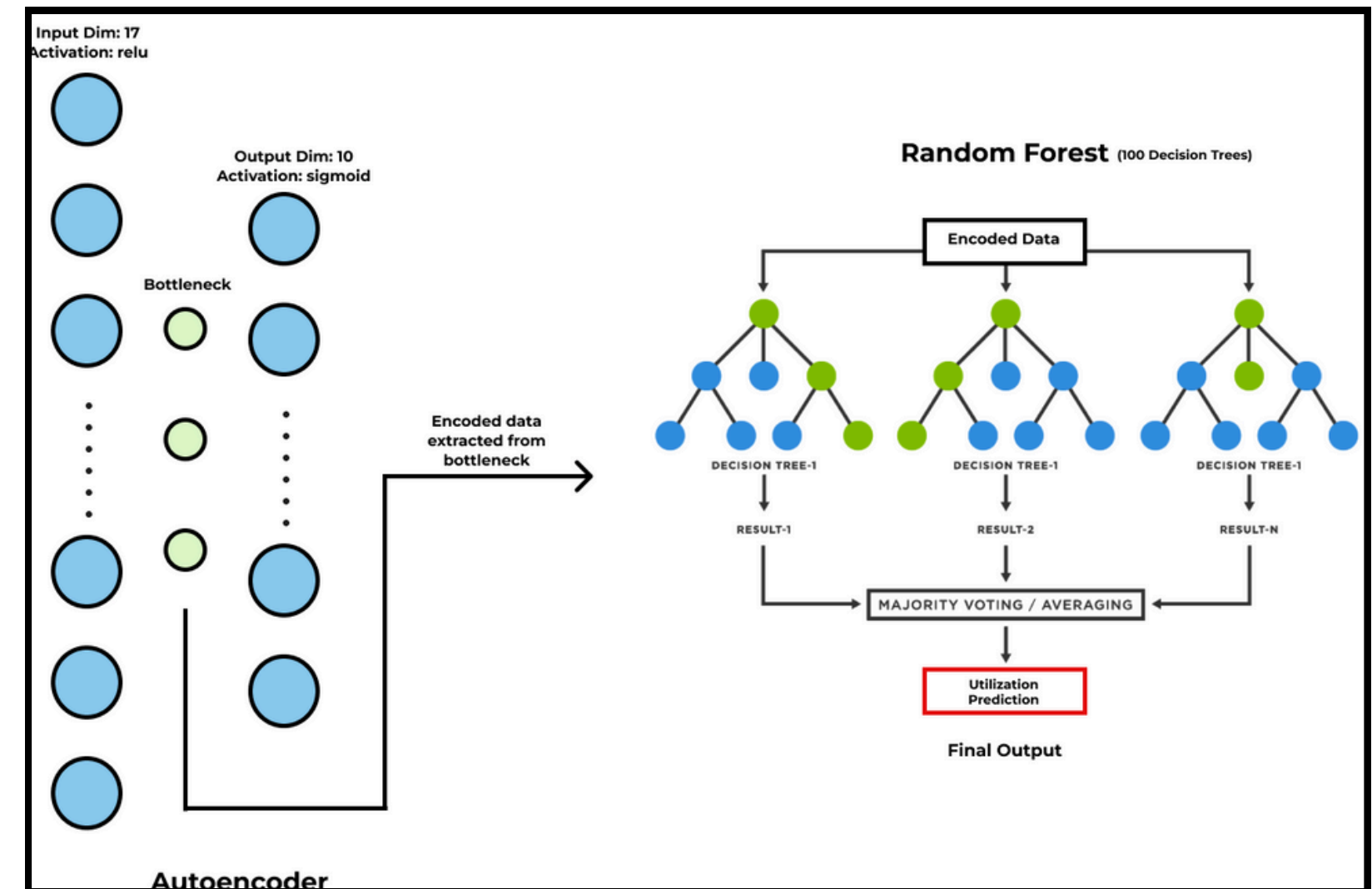
Generalization to prevent overfitting, a better understanding of features and non-linear relationships in the dataset, better predictions because of reduced dimensions

How the Model Works:

The grouped data is first segregated by strategies to prepare it for modeling separately. Separate Autoencoder models are created based on the separately grouped strategy data. The model reconstructs the input data and encodes it to a compressed version, capturing important features in a lower dimensional space. This is stored in the 'encoder_model' variable. After this step, a Random Forest predictor is trained using the encoded input set from the Autoencoder and the target variable, which is the utilization. The trained random forest predictor is tested using the encoded test set to make utilization predictions.

Performance:

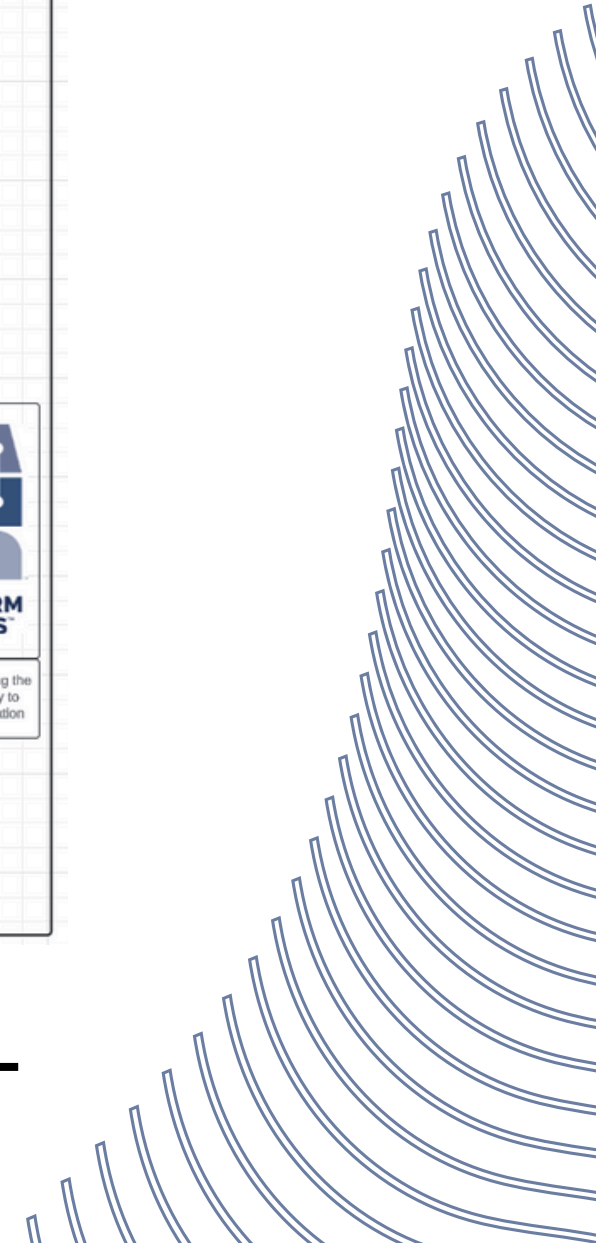
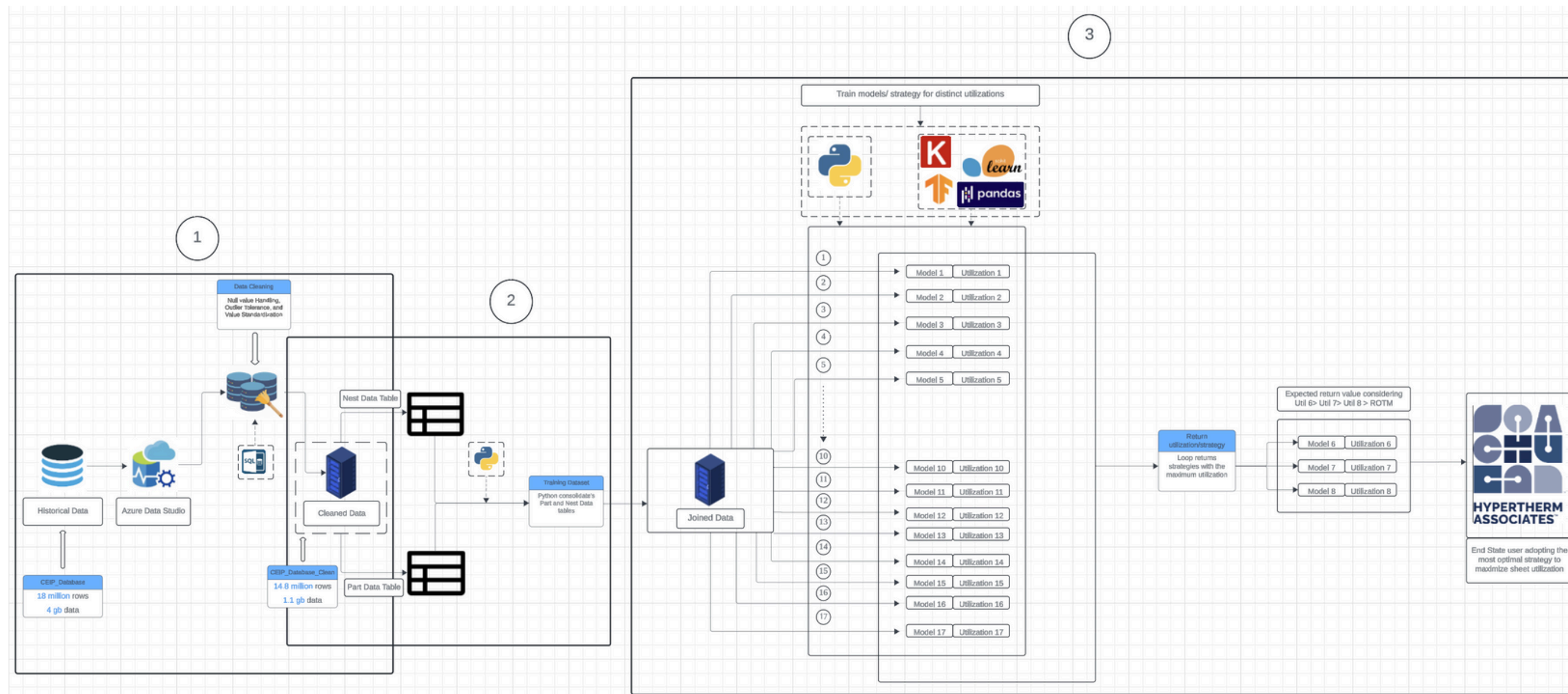
The average MAE for the AE-RF models of all the strategies averages out to 6.15%.



Result and Impact

Our Final Solution Architecture

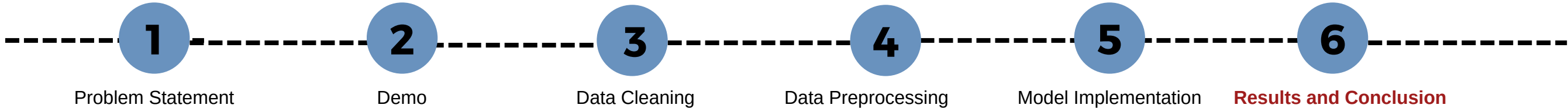
In the first step, data cleaning is performed to eliminate any erroneous data, ensuring the dataset's reliability. The second step involves data preprocessing, where relevant tables are combined, and the data is prepared for analysis, making it suitable for model training. Finally, in the third step, models are run, employing separate models for each strategy to identify and provide the top three strategies with the highest utilization.



We are suggesting the AE-RF model as our proposed solution for this use case

- The Autoencoder-Random Forest Ensemble model performed the best for predicting the utilization of Jobs, with a mean absolute error average of ~6%
- If the model is employed in Hypertherm’s nesting workflow, there is an average increase of ~10% in the utilization that people get out of cutting steel sheets

	fStrategies	CropUtil	Best Strategy Predicted	Predicted CropUtil
0	0	0.022530	64	0.662942
2	8	0.636089	128	0.739873
3	8	0.634186	16384	0.709342
4	0	0.768649	256	0.769121
5	128	0.413703	256	0.606014
...
1790931	0	0.262695	4	0.702255
1790932	4	0.640358	2048	0.811541
1790933	8	0.617906	64	0.819938
1790934	8	0.509921	64	0.739062
1790935	0	0.648776	128	0.826359



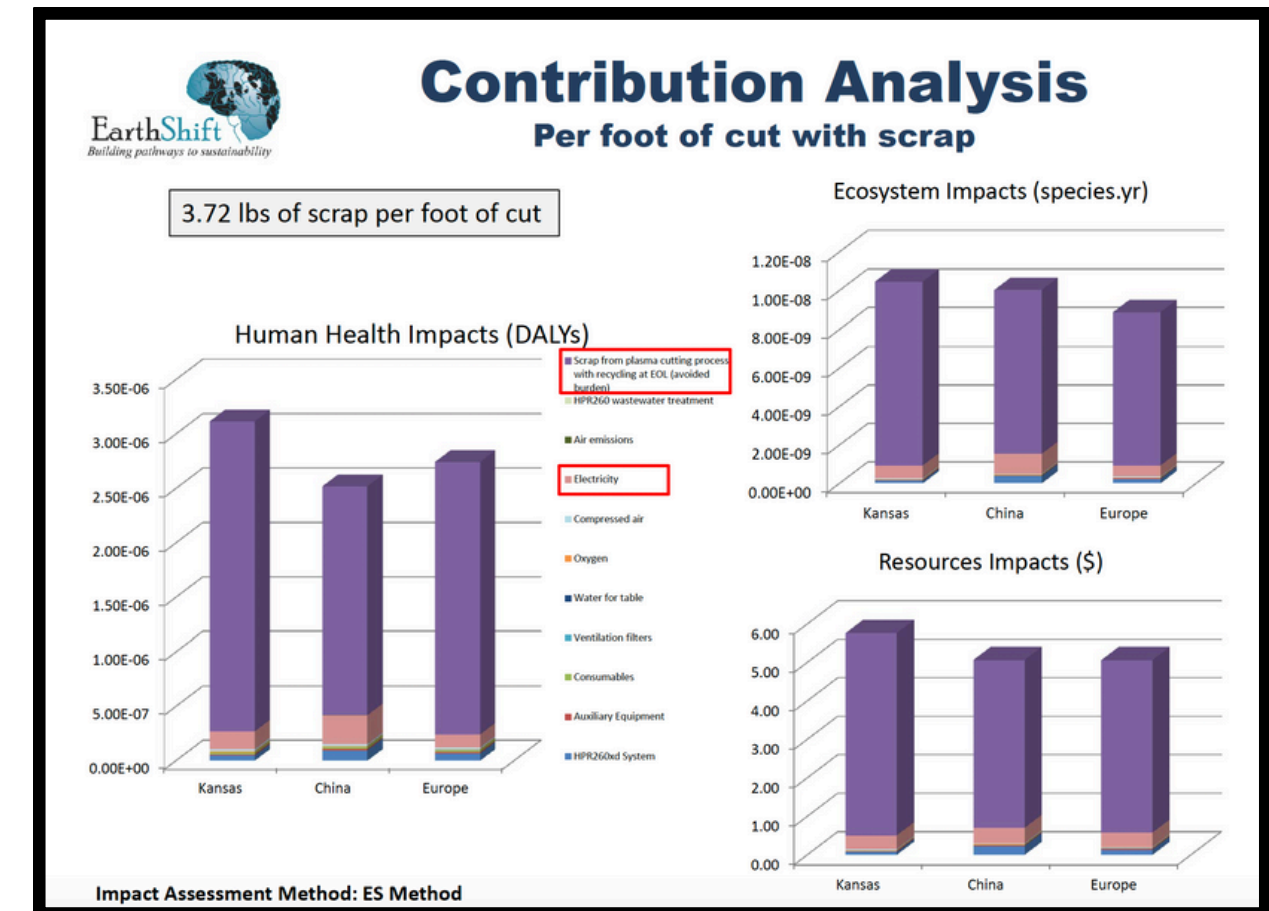
Integrating our solution into Hypertherm's approach would result in **annual savings of \$181,200**, prevent the loss of **23 days of a person's life each year**, and preserve **0.096 additional species in the area over the next 50 years**.

Assumptions:

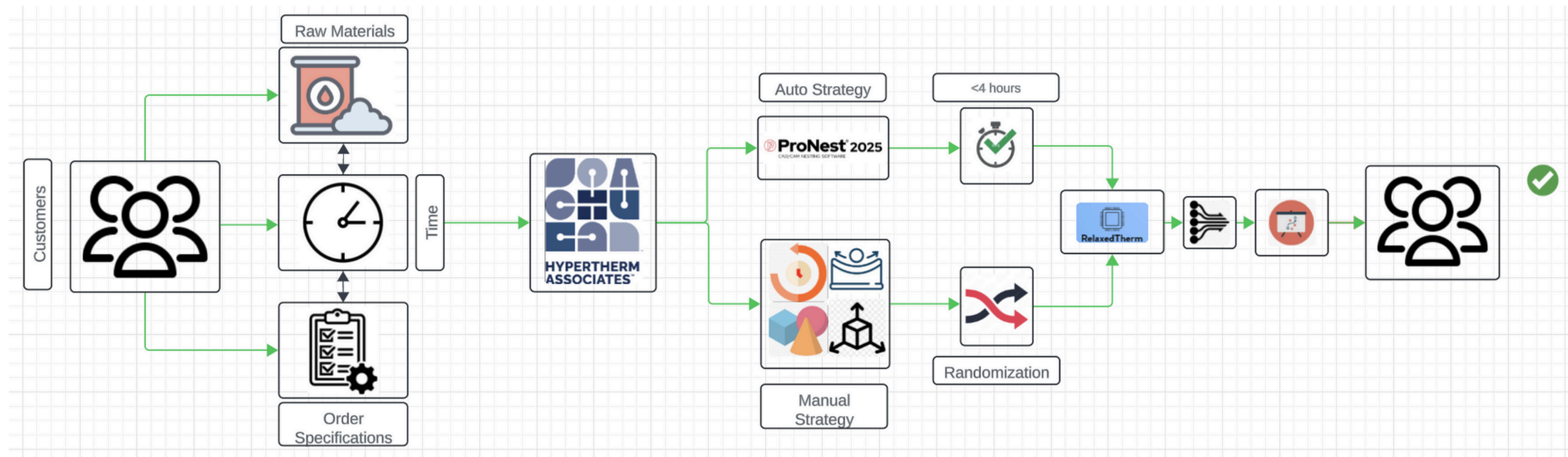
- Each part order is one foot long and has one nest.
- Only one person is working on all the orders for the entire year.
- To make the calculations streamlined we assume 10% point worth of positive impact on the metric we are considering for the calculations.

Factors Considered:

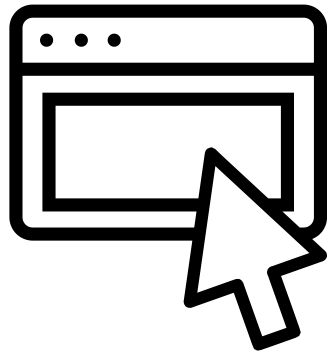
Number of Orders/year	Number of Days Saved
Feet of metal Cut/ year	Number of species that may disappear/year/order due to the impact
lbs of scrap/foot of cut	Number of species that may disappear/year due to the impact
Current Utilization	Number of species that may disappear in the next 50 years to the impact
Years cut short (DALY)	Number of species that would be saved in the next 50 years
Days cut short/ year	Number of Days Saved
Resources(\$) impacted/order/year	Resources(\$) Saved/ year



We solved the identified pain points in Hypertherm's workflow: time consumed by the software and non-optimal strategy selection, both of which significantly contributed to low raw material utilization.



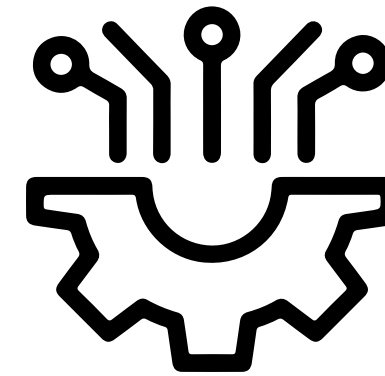
Future Scope



Integrating the Model into a
Website



Scaling the model to train on
all the part inputs



Testing out other models that
might better fit our use case

Thank you!

Thank you to **Marc and Robin** at Hypertherm for all their help on this project throughout the term!

To our Mentors: Prof. Parker, Dr. Raymond, and Ben, thanks for the incredible opportunity.

Questions?



Appendix

Calculations for Impact slide:

	Human Health Impact		Ecosystem impact		Resources impact (\$/yr)	
	Pre implementation	Post Implementation	Pre implementation	Post Implementation	Pre implementation	Post Implementation
Number of Orders/year	302,000	302,000	302,000	302,000	302,000	302,000
Feet of metal Cut/ year	302,000	302,000	302,000	302,000	302,000	302,000
lbs of scrap/foot of cut	3.72	3.426315789	3.72	3.426315789	3.72	3.426315789
Current Utilization	62.00%	65.00%	62.00%	65.00%	62.00%	65.00%
Years cut short (DALY)	0.00000209	0.000001881				
Days cut short/ year	230.3807	207.34263				
Number of Days Saved	23.03807					
Number of species that may disappear/year/order due to the impact			0.00000001	0.000000009		
Number of species that may disappear/year due to the impact			0.0112344	0.009312726316		
Number of species that may disappear in the next 50 years to the impact			0.56172	0.4656363158		
Number of species that would be saved in the next 50 years			0.09608368421			
Number of Days Saved						
Resources(\$) impacted/order/year					6	5.40
Resources(\$) impacted/year					1812000	1630800
Resources(\$) Saved/ year					181200	

MSE - All Models

MSE - Logistic Regression

Strategy -2147483648 - LR MSE: 0.0980084430462847
Strategy 0 - LR MSE: 0.0450198937374988
Strategy 1 - LR MSE: 0.026879336902438867
Strategy 2 - LR MSE: 0.011879141313046576
Strategy 4 - LR MSE: 0.02232421668214772
Strategy 8 - LR MSE: 0.02232806198457016
Strategy 16 - LR MSE: 0.018775997365779035
Strategy 32 - LR MSE: 0.021725520313180292
Strategy 64 - LR MSE: 0.020679917467253815
Strategy 128 - LR MSE: 0.02127611297982527
Strategy 256 - LR MSE: 0.01316128246361642
Strategy 512 - LR MSE: 0.023417447469896397
Strategy 1024 - LR MSE: 0.02499253496631658
Strategy 2048 - LR MSE: 0.026662445326147848
Strategy 4096 - LR MSE: 0.02580255400339677
Strategy 8192 - LR MSE: 0.02814165436948609
Strategy 16384 - LR MSE: 0.02275297703850721

MSE - ANN

Strategy -2147483648 - ANN MAE: 0.17141690402405493
Strategy 0 - ANN MAE: 0.02514191891341782
Strategy 1 - ANN MAE: 0.013655793369293621
Strategy 2 - ANN MAE: 0.005907572402820202
Strategy 4 - ANN MAE: 0.015734189601127126
Strategy 8 - ANN MAE: 0.010604672261677443
Strategy 16 - ANN MAE: 0.01066016896679092
Strategy 32 - ANN MAE: 0.01514540548180336
Strategy 64 - ANN MAE: 0.01600299332727579
Strategy 128 - ANN MAE: 0.009791450454960068
Strategy 256 - ANN MAE: 0.012629927302528729
Strategy 512 - ANN MAE: 0.014408496732857616
Strategy 1024 - ANN MAE: 0.013666928816760916
Strategy 2048 - ANN MAE: 0.010932017264043258
Strategy 4096 - ANN MAE: 0.013070337556535267
Strategy 8192 - ANN MAE: 0.012215803208491714
Strategy 16384 - ANN MAE: 0.0106639314554052

MSE - Autoencoder Random Forest Ensemble

Strategy -2147483648 - Random Forest MSE: 0.021209262390340572
Strategy 0 - Random Forest MSE: 0.01712972512382457
Strategy 1 - Random Forest MSE: 0.011136618978461396
Strategy 2 - Random Forest MSE: 0.003770892723788292
Strategy 4 - Random Forest MSE: 0.01097673187574006
Strategy 8 - Random Forest MSE: 0.008161786868054641
Strategy 16 - Random Forest MSE: 0.009344964244166013
Strategy 32 - Random Forest MSE: 0.010108515820289005
Strategy 64 - Random Forest MSE: 0.012732688856081968
Strategy 128 - Random Forest MSE: 0.008591365865944955
Strategy 256 - Random Forest MSE: 0.006884096423003518
Strategy 512 - Random Forest MSE: 0.010699315398865959
Strategy 1024 - Random Forest MSE: 0.010100297185384794
Strategy 2048 - Random Forest MSE: 0.00784009505282407
Strategy 4096 - Random Forest MSE: 0.010265539462260964
Strategy 8192 - Random Forest MSE: 0.012707406304693332
Strategy 16384 - Random Forest MSE: 0.0087109323054738

MAE - All Models

MAE - LR

Strategy -2147483648 - LR MAE: 0.1724384731017937
Strategy 0 - LR MAE: 0.17007085061048358
Strategy 1 - LR MAE: 0.12087709037506887
Strategy 2 - LR MAE: 0.07587781897573141
Strategy 4 - LR MAE: 0.10831671082297097
Strategy 8 - LR MAE: 0.11130536659789568
Strategy 16 - LR MAE: 0.10089318489413998
Strategy 32 - LR MAE: 0.11463965895750547
Strategy 64 - LR MAE: 0.10674503232528745
Strategy 128 - LR MAE: 0.11334468477553149
Strategy 256 - LR MAE: 0.09056337047297233
Strategy 512 - LR MAE: 0.11708796531329728
Strategy 1024 - LR MAE: 0.12435057973973052
Strategy 2048 - LR MAE: 0.12670126339086446
Strategy 4096 - LR MAE: 0.12620449579268
Strategy 8192 - LR MAE: 0.1336700162174162
Strategy 16384 - LR MAE: 0.1184001536864068

MAE - ANN

Strategy -2147483648 - ANN MAE: 0.1097107682471816
Strategy 0 - ANN MAE: 0.1004535755441663
Strategy 1 - ANN MAE: 0.0740200648623339
Strategy 2 - ANN MAE: 0.03738127748469544
Strategy 4 - ANN MAE: 0.07874878821943923
Strategy 8 - ANN MAE: 0.06088854574044074
Strategy 16 - ANN MAE: 0.06740776591310316
Strategy 32 - ANN MAE: 0.08809113846669805
Strategy 64 - ANN MAE: 0.09188961986987759
Strategy 128 - ANN MAE: 0.07266957436902598
Strategy 256 - ANN MAE: 0.08451832628498428
Strategy 512 - ANN MAE: 0.08299802368316325
Strategy 1024 - ANN MAE: 0.07359543907081563
Strategy 2048 - ANN MAE: 0.06385327250228227
Strategy 4096 - ANN MAE: 0.07603518800236708
Strategy 8192 - ANN MAE: 0.07218181656478946
Strategy 16384 - ANN MAE: 0.07263756478868126

MAE - Autoencoder Random Forest Ensemble

Strategy -2147483648 - Random Forest MAE: 0.09305657531236997
Strategy 0 - Random Forest MAE: 0.07994441079713896
Strategy 1 - Random Forest MAE: 0.0627151977764201
Strategy 2 - Random Forest MAE: 0.027509728272255507
Strategy 4 - Random Forest MAE: 0.0616838621334025
Strategy 8 - Random Forest MAE: 0.050668070367750534
Strategy 16 - Random Forest MAE: 0.05650405574998458
Strategy 32 - Random Forest MAE: 0.06422683678323825
Strategy 64 - Random Forest MAE: 0.07012840373321899
Strategy 128 - Random Forest MAE: 0.0579891598706541
Strategy 256 - Random Forest MAE: 0.054372319788147924
Strategy 512 - Random Forest MAE: 0.06566009398828265
Strategy 1024 - Random Forest MAE: 0.05673849683360188
Strategy 2048 - Random Forest MAE: 0.04805366126191594
Strategy 4096 - Random Forest MAE: 0.06200144543337612
Strategy 8192 - Random Forest MAE: 0.07299470779275806
Strategy 16384 - Random Forest MAE: 0.060406469730801014

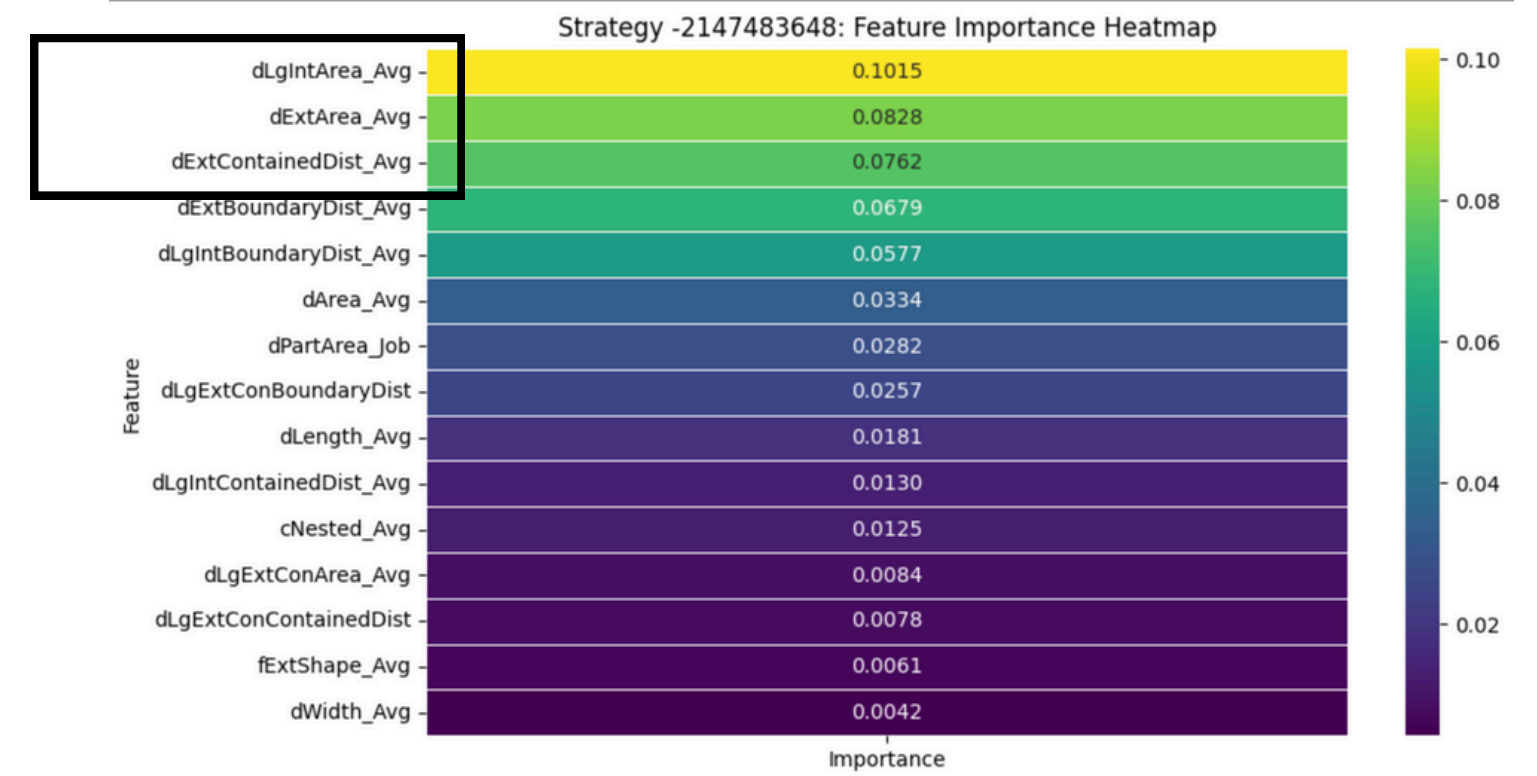
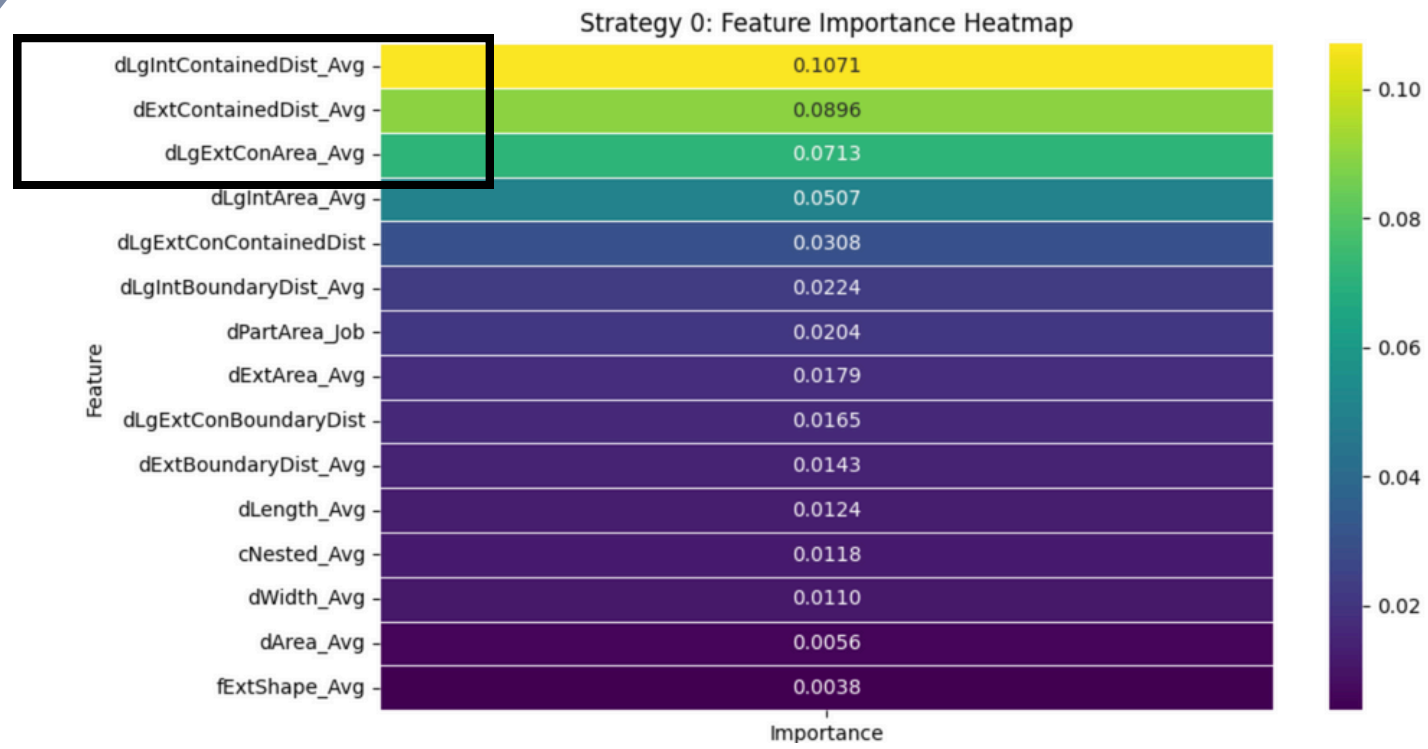
Metrics used to calculate impact

Impact Assessment Method

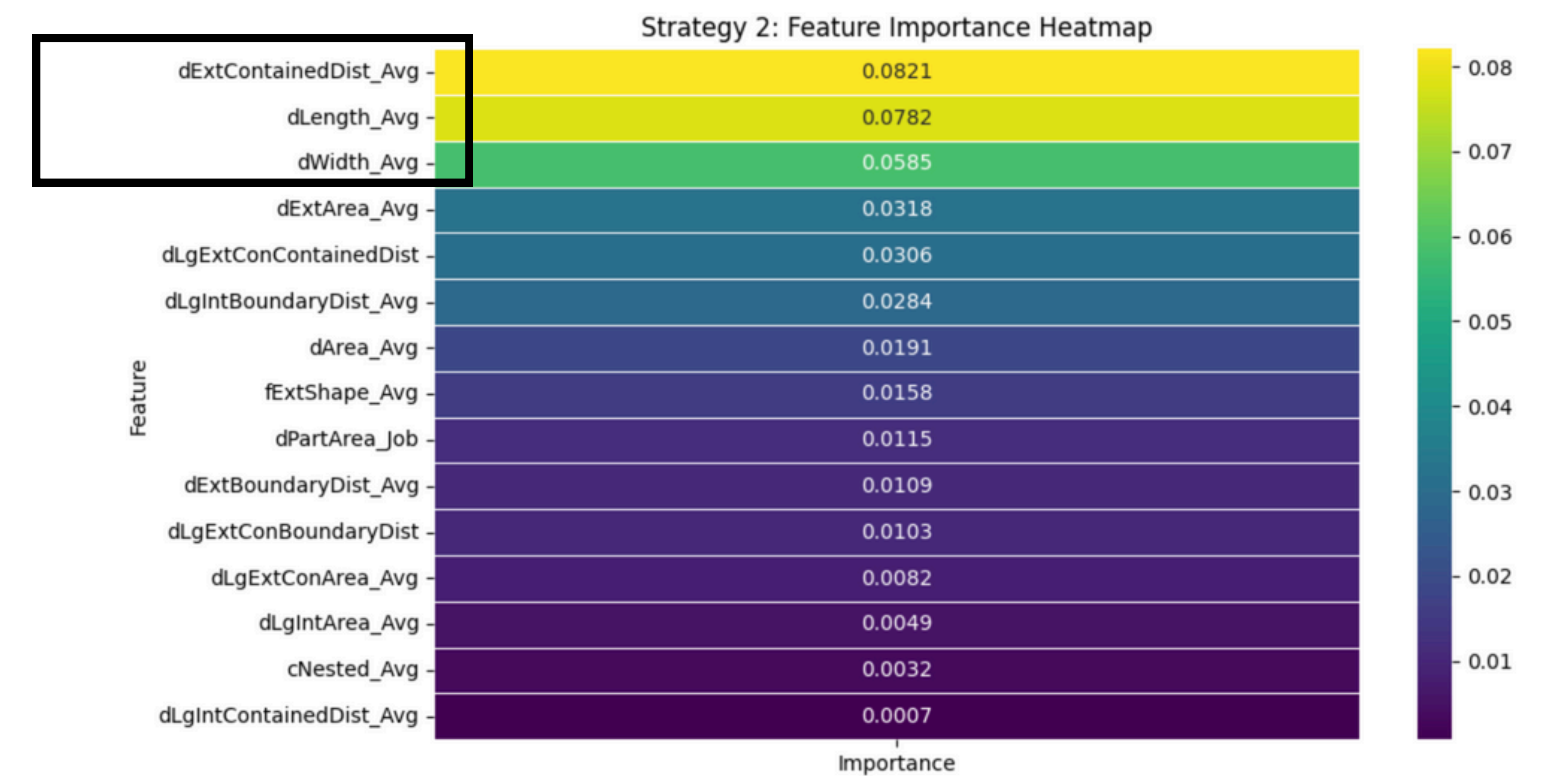
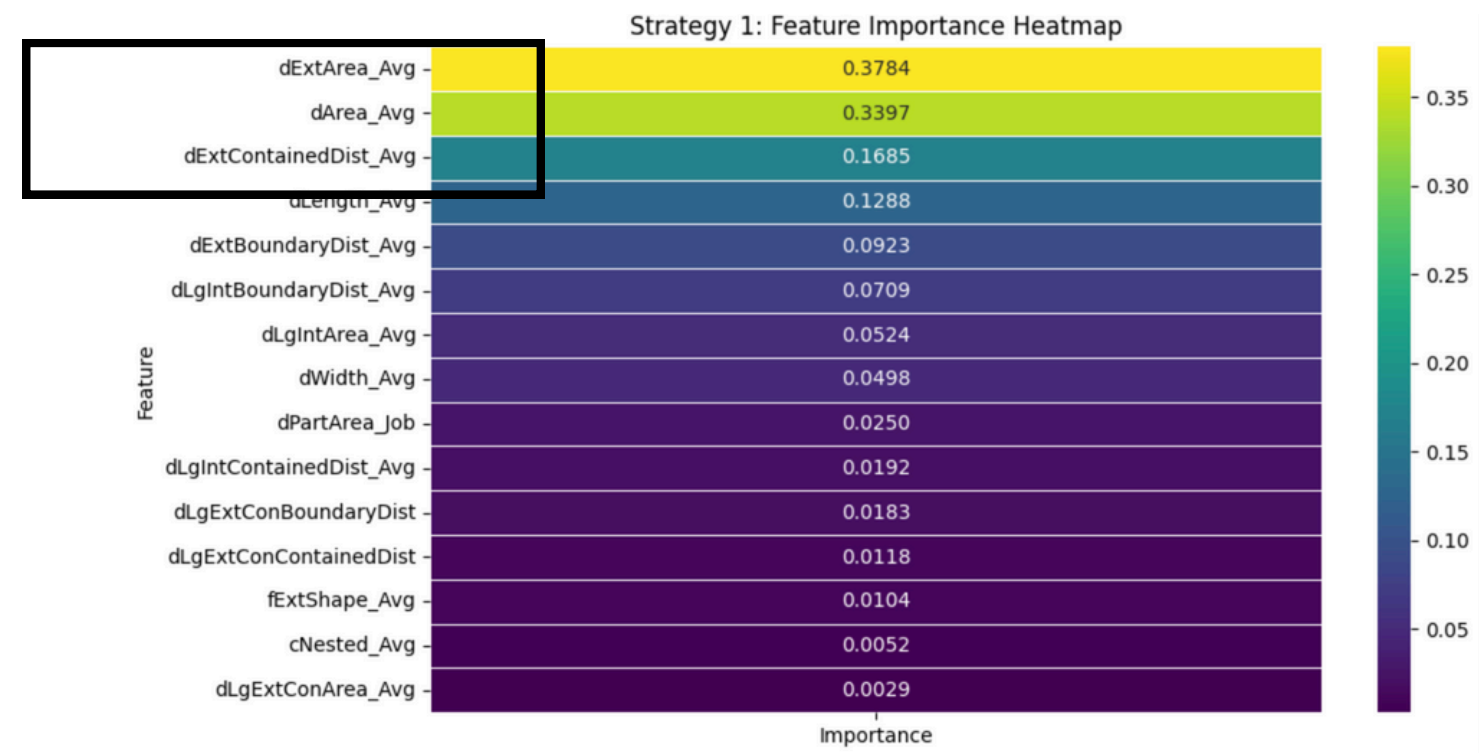
Damage Category	Units	From	Comments
Human health	DALY	ReCiPe Endpoint (H)	Accounts for years lived disabled as well as life cut short
Ecosystems	Species * yr	ReCiPe Endpoint (H)	Assessed in units of species * yr, or the number of species that may disappear due to the impact times the area over which they are affected times the duration that the species are affected
Resources	Economic units	ReCiPe Endpoint (H)	Puts a future value on resources which will be unavailable since we are using them today
Climate change	kg CO ₂ eq.	IPCC 100a	Same method used by most GHG accounting programs
Water	m ³	ReCiPe Midpoint (H)	Counts the amount of water consumed. Does not show impact. Used for benchmarking only.
CED	MJ	Cumulative Energy Demand	Adds up different categories of energy

Impact Assessment Method: ES Method





Every strategy has unique correlations with various features, which is why we've incorporated all these features and it is also why we have strategy specific models.



Libraries used

NumPy: For numerical computations and array manipulations.

Pandas: For data manipulation and analysis.

scikit-learn:

‘**model_selection**’ for splitting datasets into training and testing sets.

‘**preprocessing**’ for feature scaling.

‘**ensemble**’ for regression tasks using random forests.

‘**metrics**’ for evaluating model performance.

‘**linear_model**’ for linear regression models.

TensorFlow/Keras:

‘**models**’ for building and training neural networks.

‘**layers**’ for adding dense and dropout layers.

‘**callbacks**’ for improving model training with early stopping and learning rate reduction.

‘**optimizers**’ for optimizing neural networks using Adam.randomization of strategy