MONTANA
Department of Transportation

ARTIFICIAL INTELLIGENCE (AI) DRIVEN TOOL FOR PROJECT DURATION ESTIMATION

Final Presentation

May 23rd, 2023

David Jeong, Ph.D.
Hamid Alikhani, GRA

TEXAS A&M
UNIVERSITY

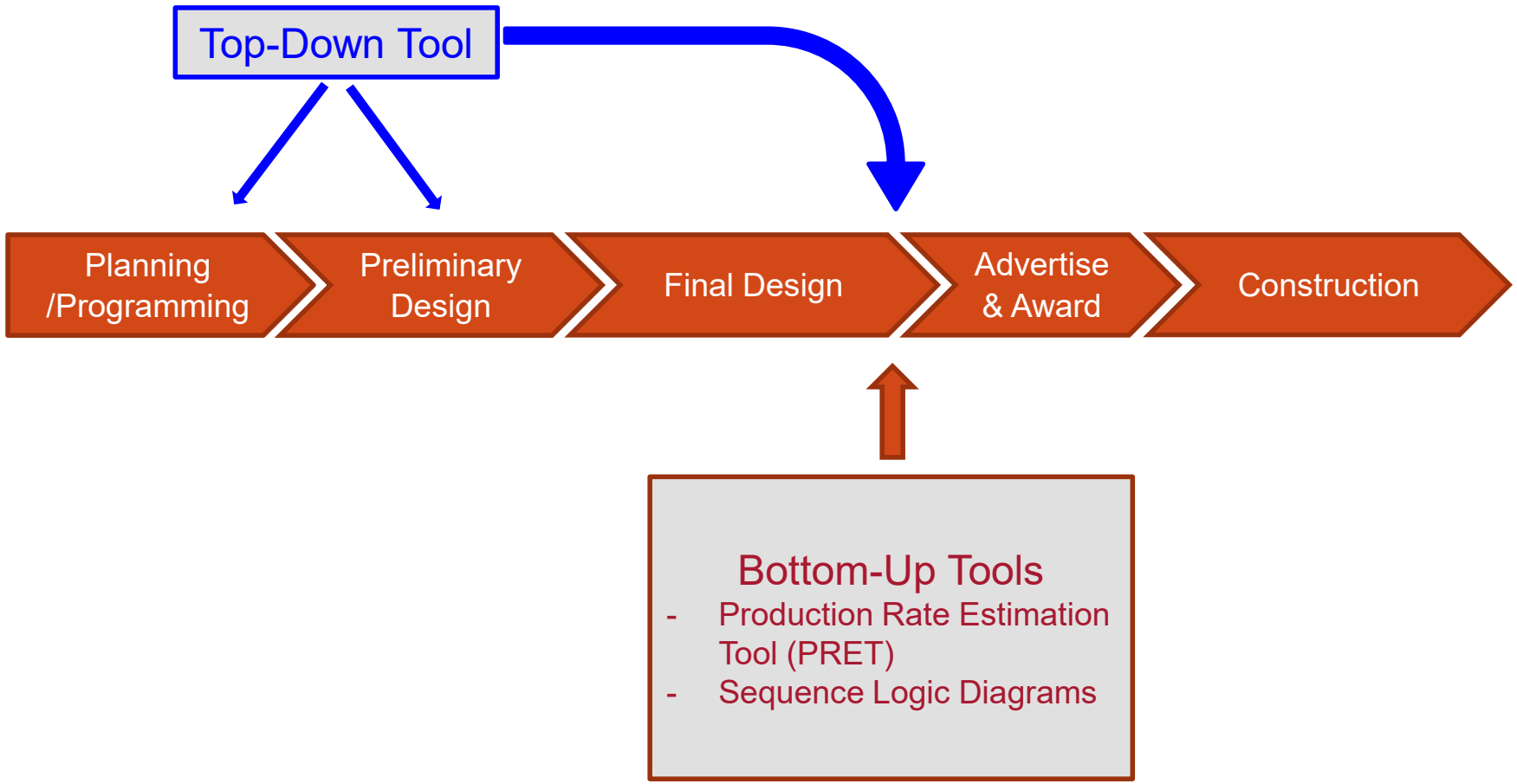
Research Goal

- Develop a **Top-Down Project Duration Estimation Tool** to estimate project duration during the pre-construction stage when a limited amount of project information is available.

Key idea:

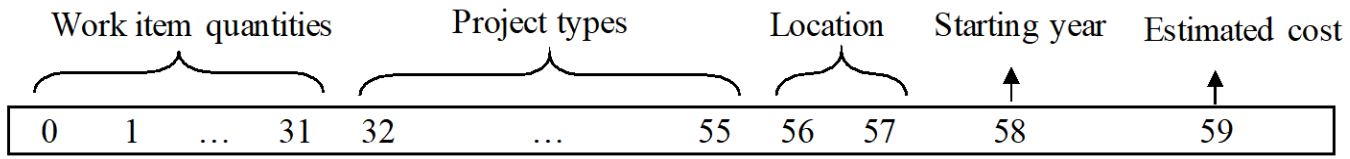
There is a strong correlation between project duration and key project characteristics such as estimated cost, and quantities of major work items.

Top-Down Tool



Data Collection

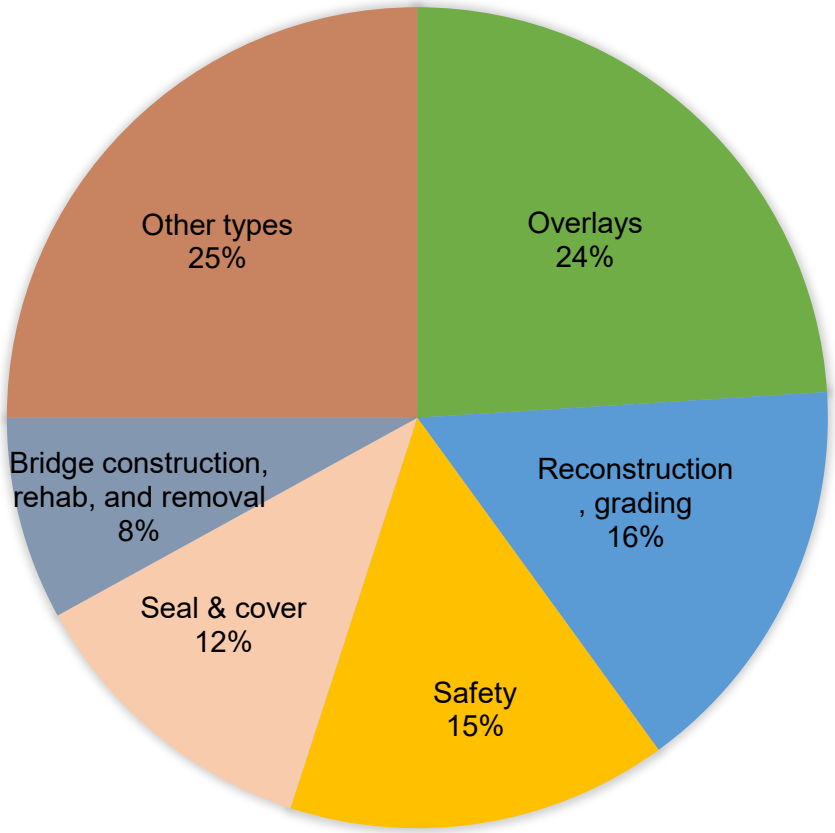
- Bid tabulation data and daily work report data of 1,090 highway projects from 2008 to 2019
- Data attributes:
  - Project #
  - Project location
  - Project type
  - Project year
  - Work item quantities
  - construction cost
  - Project duration



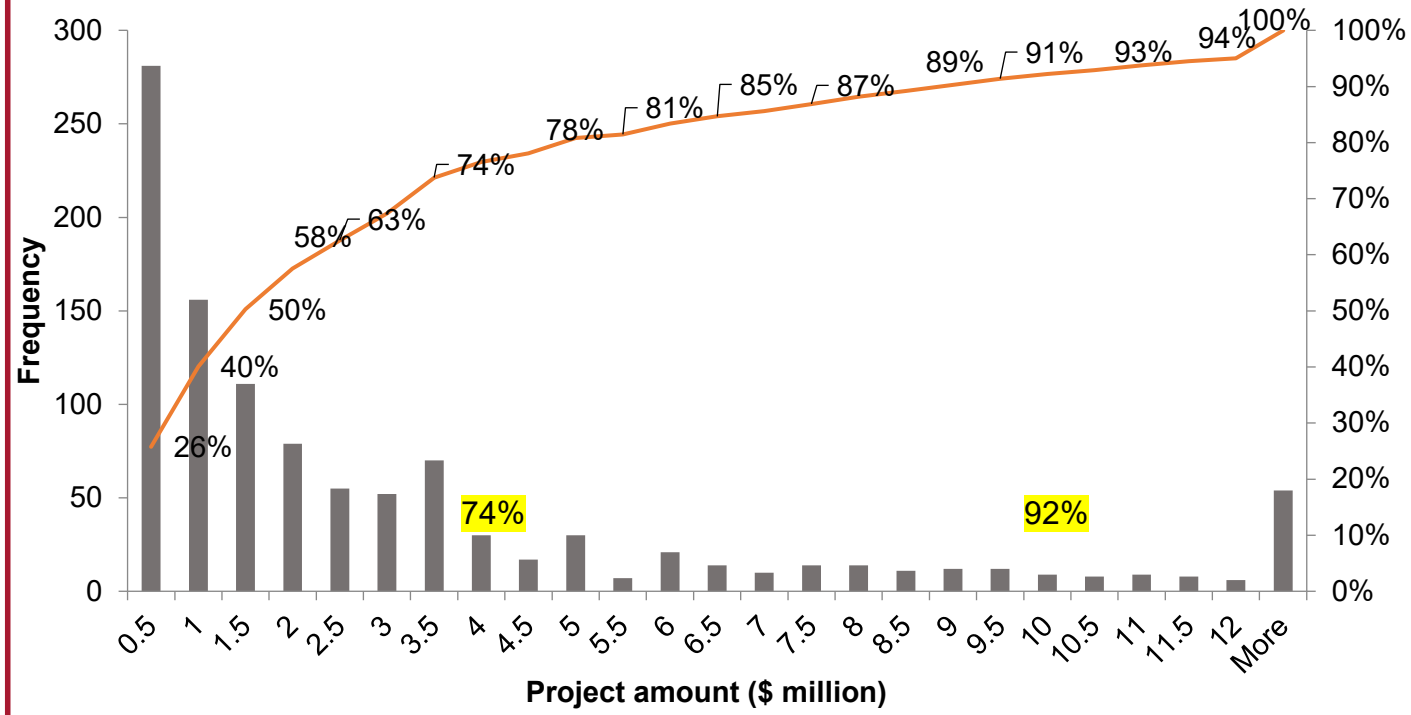
Project Attributes

Work Types

- Five most common project types account for 76% of all highway projects in MDT



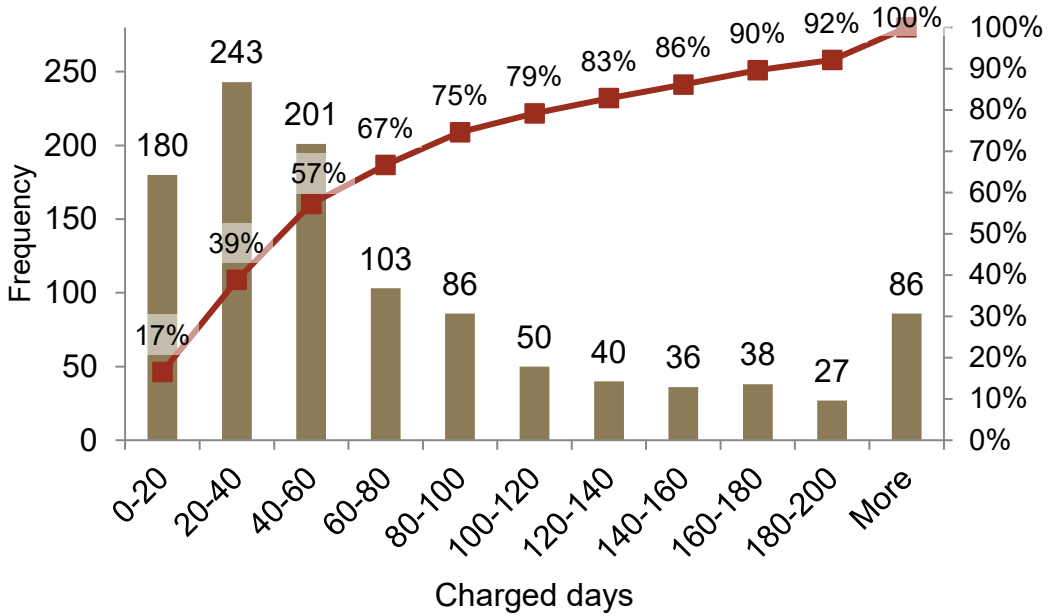
Construction Cost Distribution



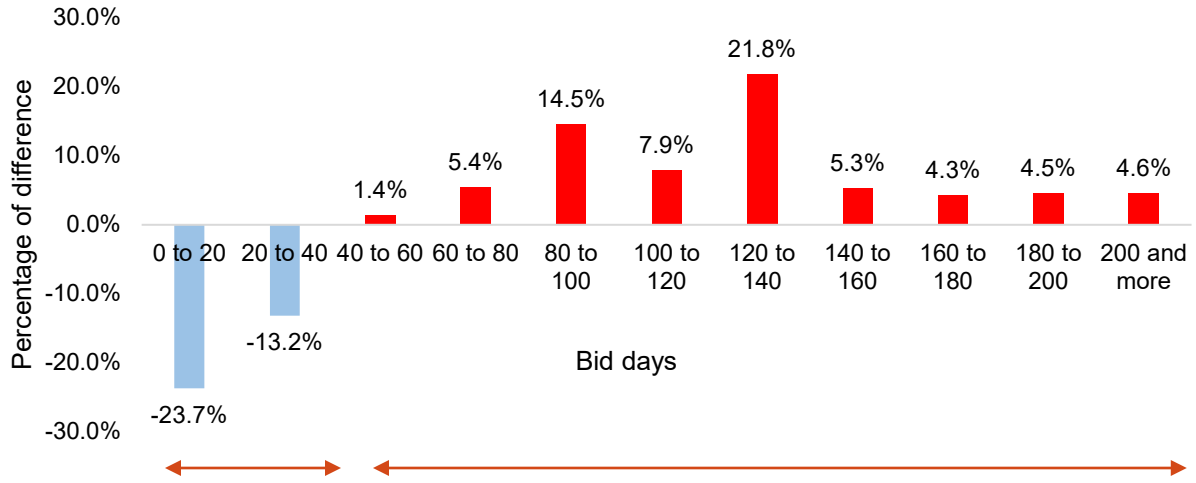
- All project costs are adjusted to the base year of 2018 using the National Highway Construction Cost Index (HCCI)
- Approximately half of the projects cost less than $2M
- 74% of projects are less than $4M
- 92% of projects are less than $10M

Project Schedule Performance

- Frequency analysis of charged days



Schedule Performance : $\frac{Charged\ days - Bid\ days}{Bid\ days} * 100$



Projects completed sooner than estimated

Projects completed later than estimated

Controlling Work items

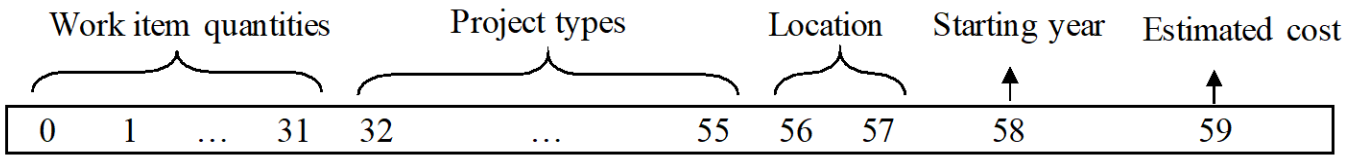
- Frequency analysis of controlling work items (from 1,090 projects)

Row	Controlling Work Item	Frequency
1	Mobilization	1090
2	Traffic control	1086
3	Remove existing structures	868
4	Pavement marking	839
5	Emulsified asphalt	718
6	Cover	700
7	Signs	697
8	Temporary activities	596
9	Crushed aggregate course	500
10	Base preparations (soil stabilization)	460
11	Guard rail	460
12	Milling and pulverizing	457
13	Asphalt cement	367
14	Plant mix surfacing	354
15	Seeding	340
16	Topsoil-salvaging and placing	321

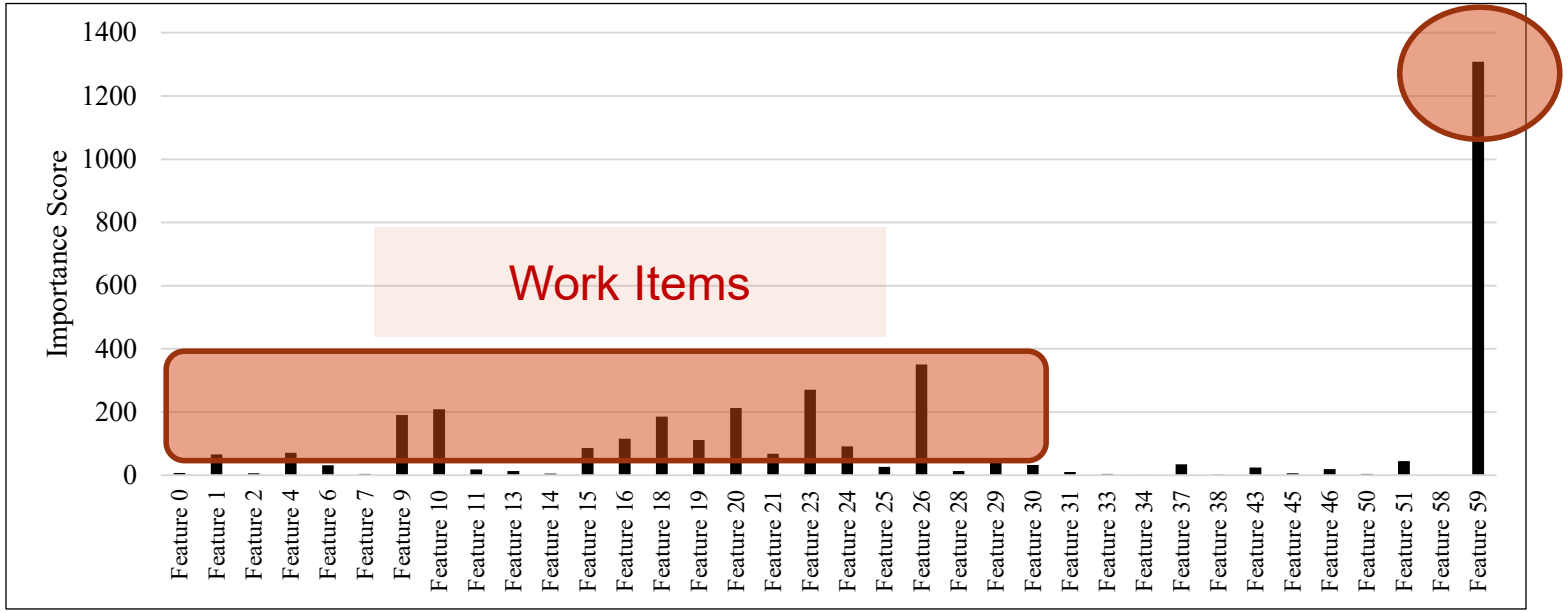
Row	Controlling Work Item	Frequency
17	Rumble strips	308
18	Commercial mix	272
19	Excavation-unclassified	272
20	Farm fence	270
21	Drainage pipe (<=24 in)	264
22	Bridge deck	246
23	Riprap	234
24	Special borrow	219
25	Sidewalk	181
26	Curb and gutter	176
27	Embankment in place	158
28	Drainage pipe (> 24 in)	155
29	Deck grooving (after curing)	125
30	Reinforcing steel	113
31	Bridge deck repair	103
32	Piling	93

Selection of Input Variables (Feature Selection)

- 59 Variables

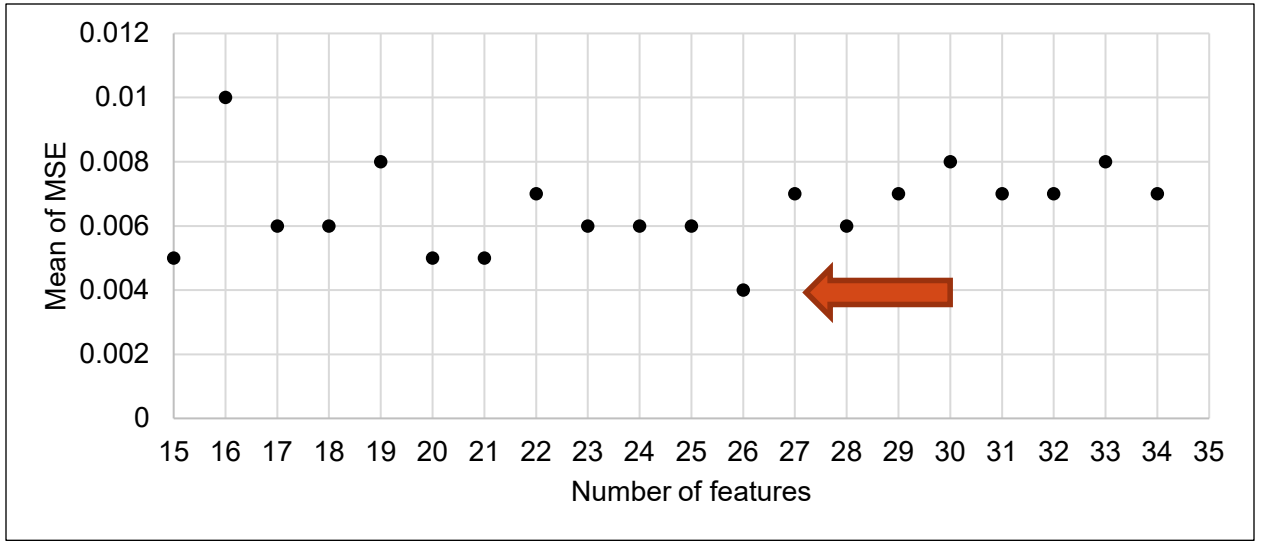


- Feature selection: Importance score of each variable in terms of project duration estimation



Selection of Input Variables (Feature Selection)

- Using all features does not guarantee the lowest MSE
- Select the best number of features: 26



- 26 features were selected including
  - Estimated construction cost (engineer's estimate in new projects)
  - 25 controlling work items

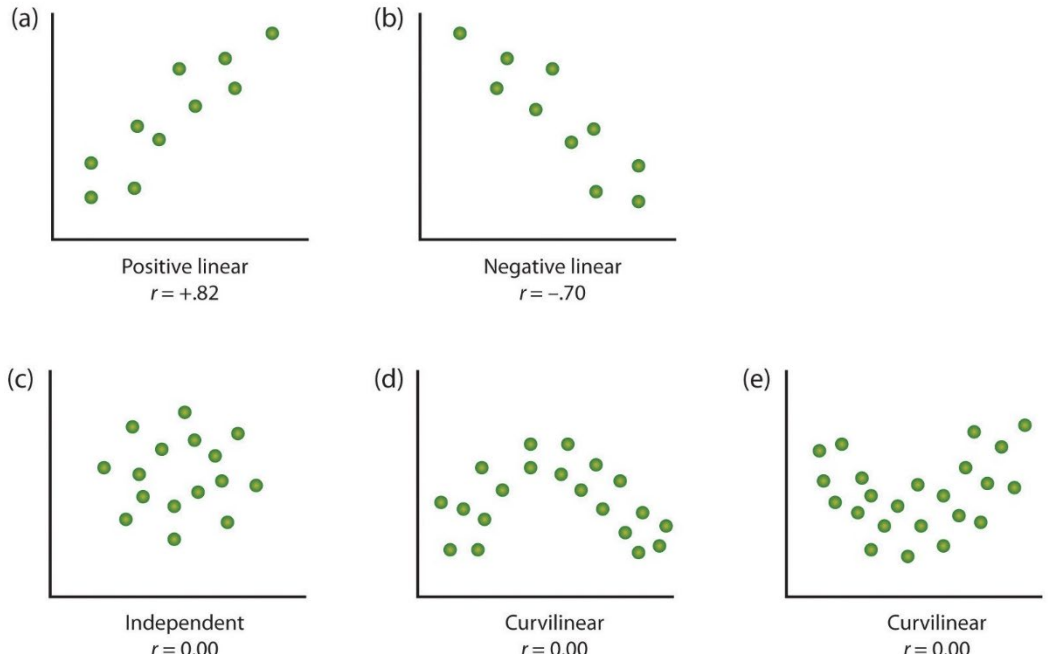
Selection of Input Variables (Feature Selection)

- 26 features were selected including
  - Estimated construction cost (engineer's estimate)
  - 25 controlling work items

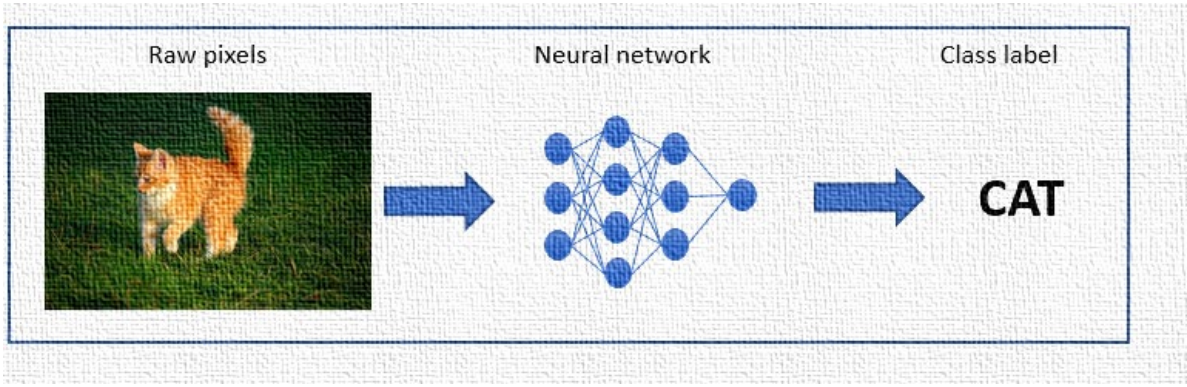
Row	Controlling Work Items	Row	Controlling Work Items
1	Traffic control	14	Piling
2	Drainage pipe (<= 24 in)	15	Sidewalk
3	Crushed aggregate course	16	Curb and gutter
4	Excavation-unclassified	17	Guard rail
5	Mobilization	18	Bridge deck
6	Seeding	19	Commercial mix
7	Special borrow	20	Signs
8	Reinforcing steel	21	Embankment in place
9	Plant mix surfacing	22	Emulsified asphalt
10	Asphalt cement	23	Cover
11	Drainage pipe (> 24 in)	24	Rumble strips
12	Deck grooving (after curing)	25	Milling and pulverizing
13	Riprap		

Benefits of Artificial Intelligence (A.I.)

- Capable of detecting complex patterns
- Especially in problems with many features



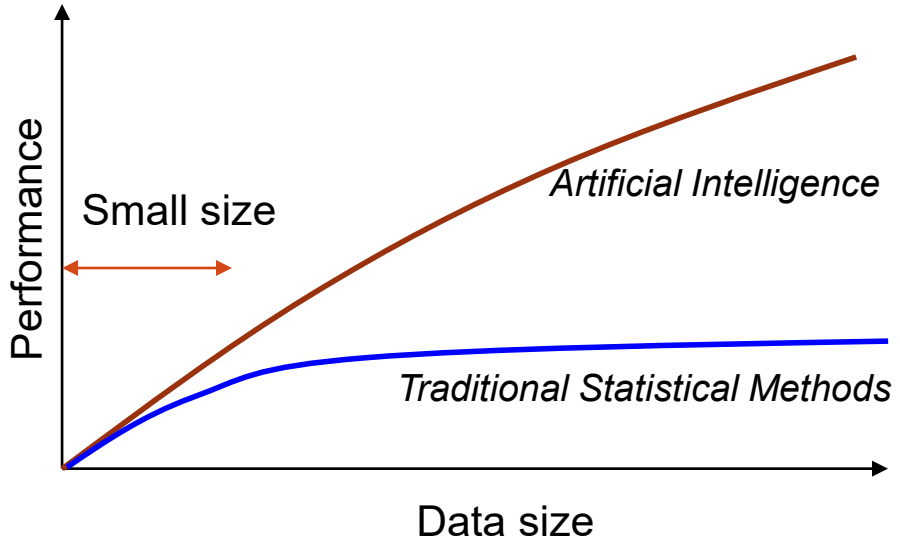
Different types of relationships between variables



AI capability in pattern recognition

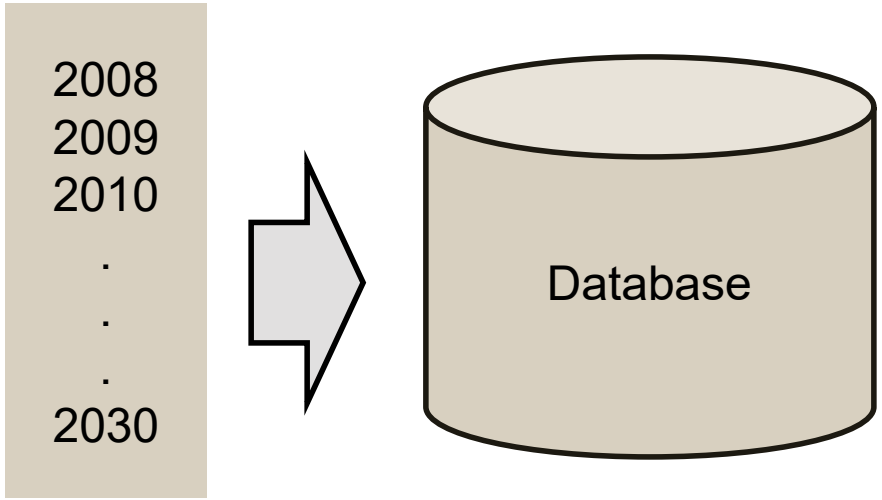
Performance of A.I. Models

- The performance of AI is generally expected to perform better as the data size gets larger while traditional statistics-based methods tend to converge to a certain point and never improve.

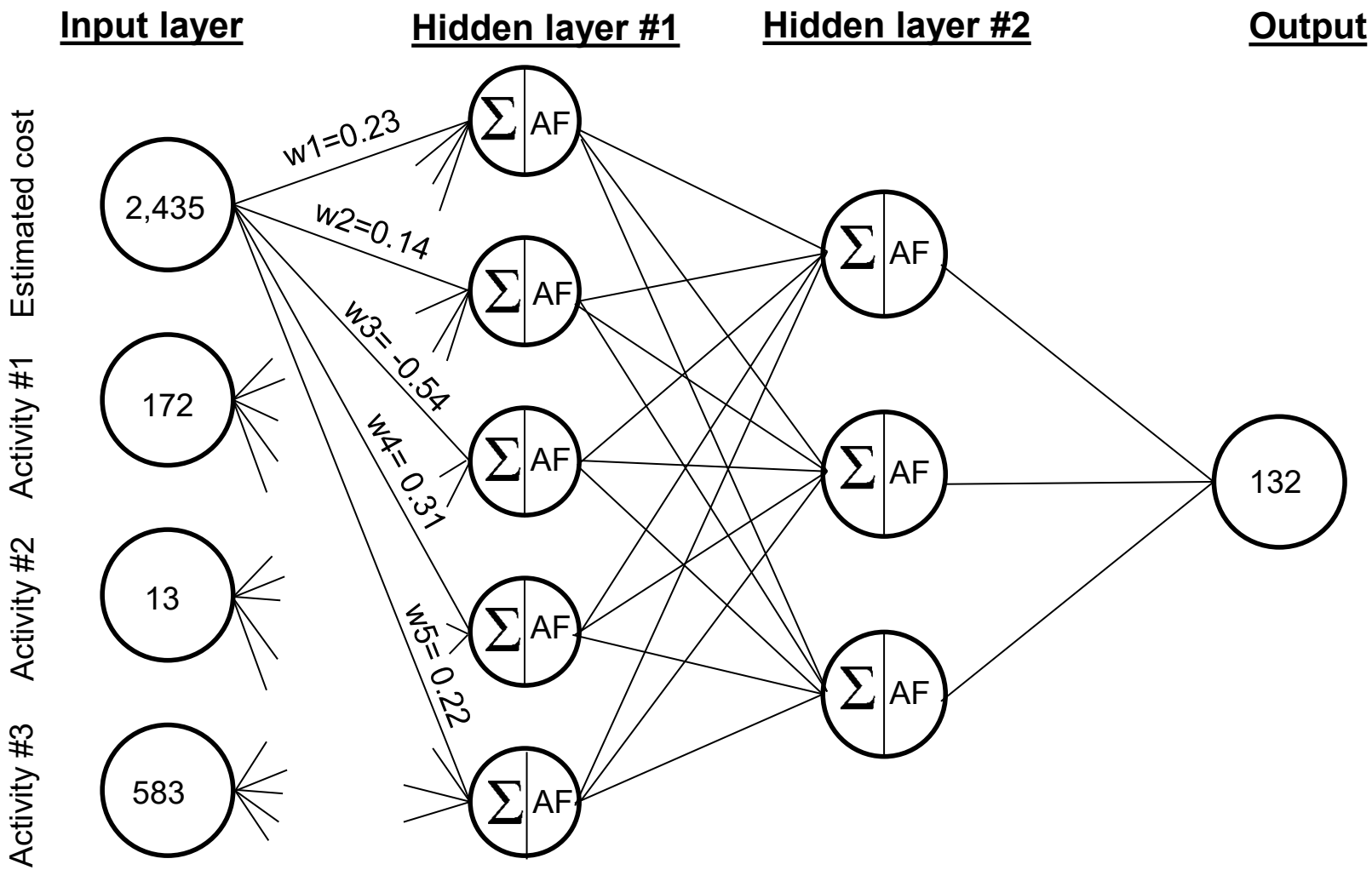


High Performance of AI in big data *(Oppermann 2019)*

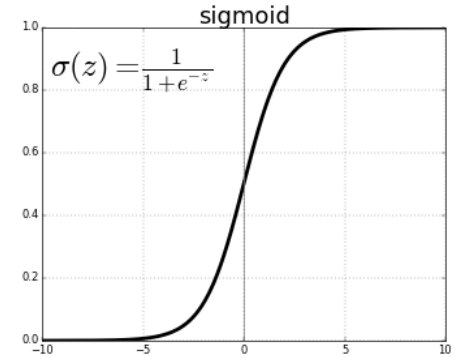
MDT's data will increate over time

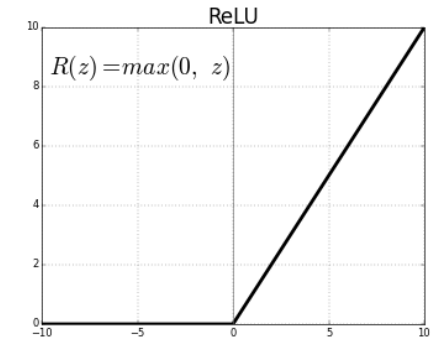


Artificial Neural Networks (ANN)



Activation Function



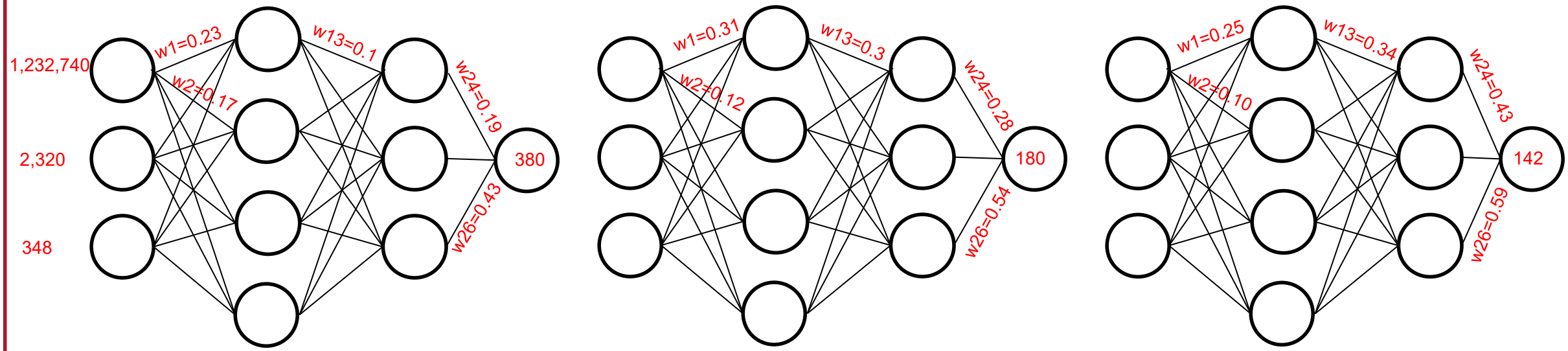


Model Training

Step 1: starting with random weights

Step 2: based on the difference between the actual output and the model's output, the weights are adjusted

Step 3: the process continues until weights are converged and the model's output gets close to the actual output

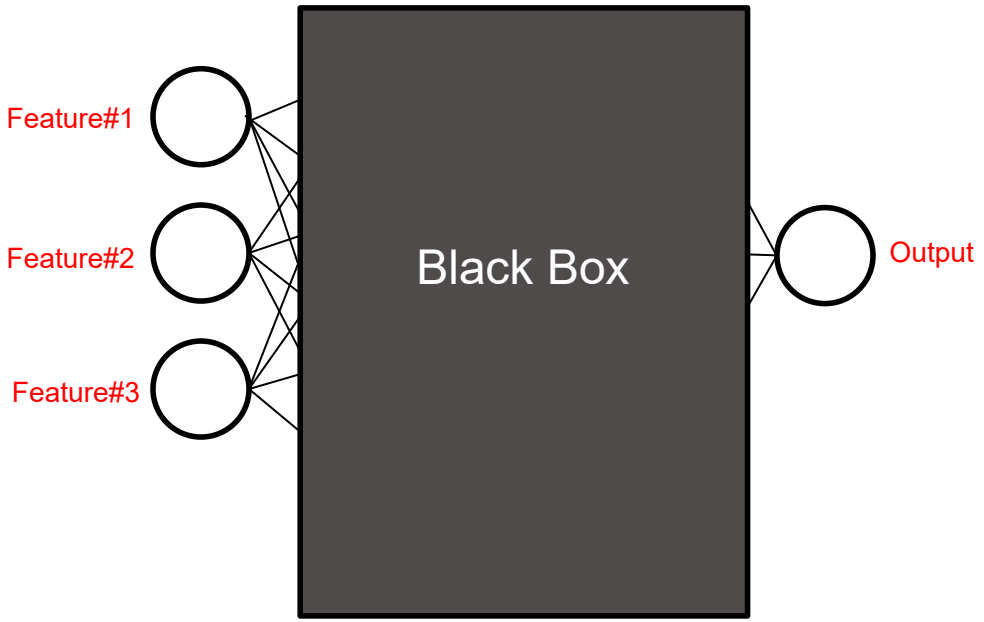


Model Training

- The goal is to determine final weights
- The ANN model's description:

Number of input features	26
Number of hidden layers	2
Number of output cells	1
Number of weights	68,000
Activation functions	Relu, Sigmoid

The ANN model receives input variables and gives the output



Model Testing

- The whole dataset is divided into 80% training dataset and 20% testing dataset.
- Test dataset is used as unseen data to evaluate the performance of the model.
- Mean Squared Error (MSE) is used as the metric for evaluation.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE = mean squared error

n = number of data points

Y_i = observed values

$\hat{Y}_i$ = predicted values

Test datapoint i

Input:

Engineer's estimate: $44,785

Activity #1 amount: 1,087

Activity #2 amount: 453

…

output:

Project duration: 37 Y_i

Model's output: 39 $\hat{Y}_i$

AI vs Regression Models

- In addition to AI model, we have developed a regression model

- Regression model:

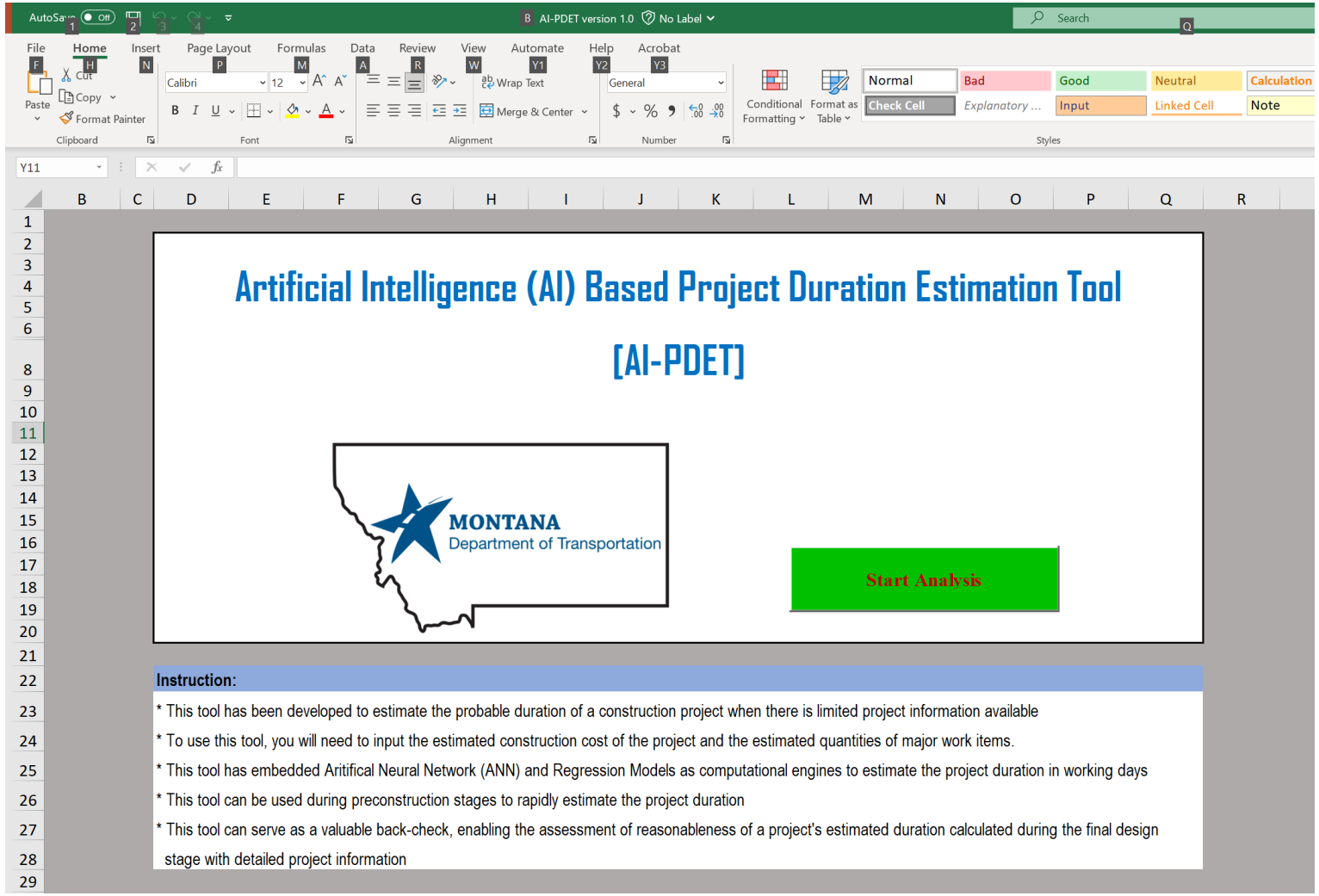
$$Y = b_0 + b_1*X_1 + b_2*X_2 + \ldots + b_n*X_n$$

X_i: i^{th} input variable, b_i : i^{th} coefficient , Y: output

Final model comparison

Model	MSE	R-squared
ANN	0.0022	0.72
Linear Regression	0.0034	0.75

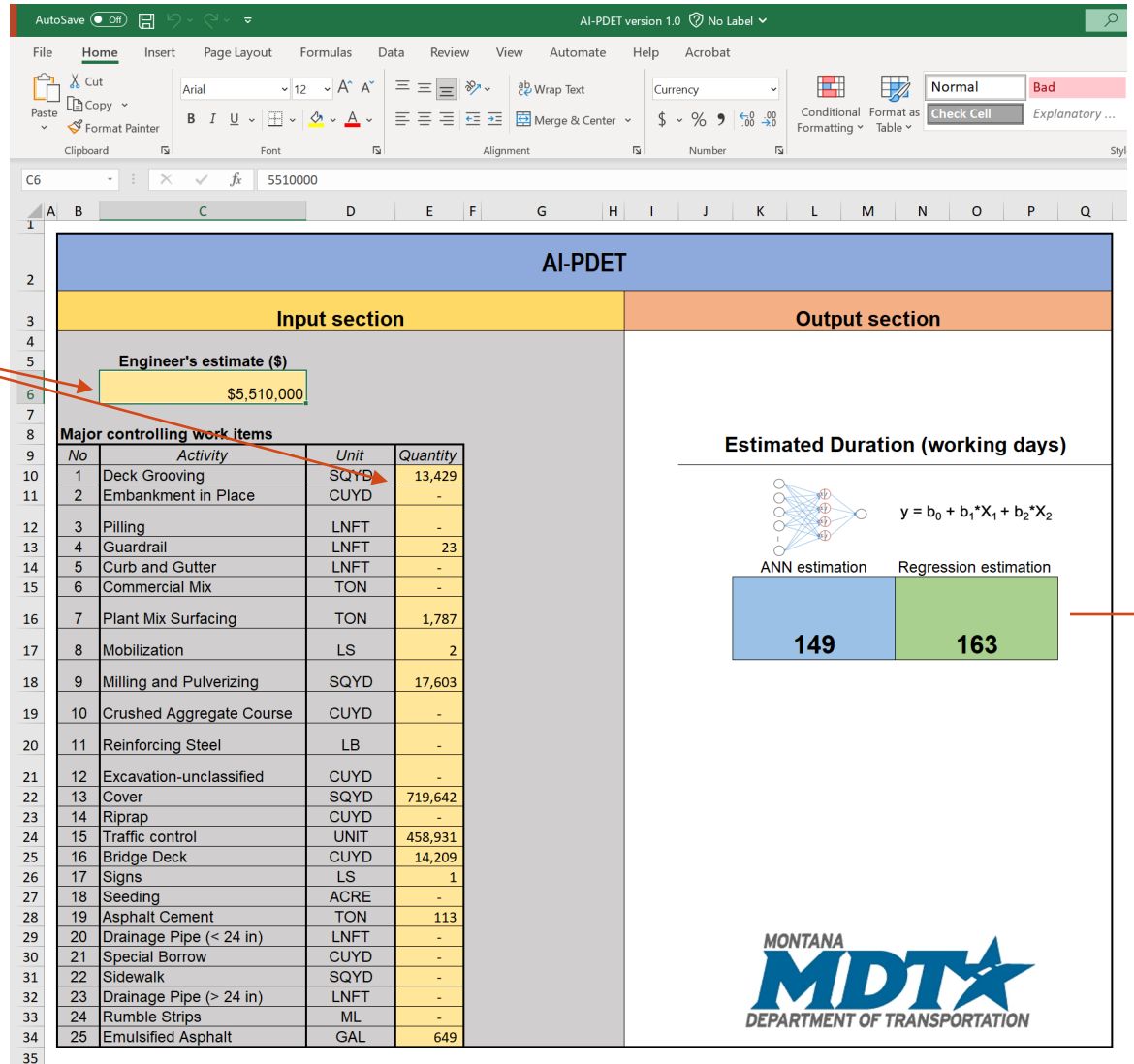
MS Excel Tool : AI-PDET



General instructions

MS Excel Tool: AI-PDET

Yellow cells are input variables that must be entered by the user

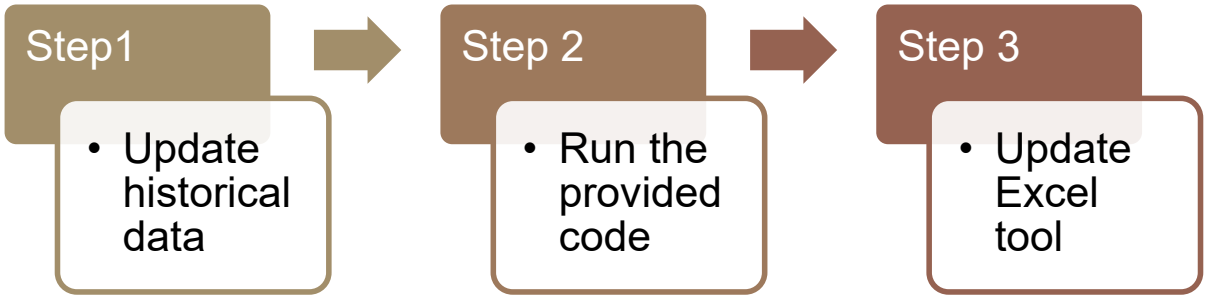


No	Activity	Unit	Quantity
1	Deck Grooving	SQYD	13,429
2	Embankment in Place	CUYD	-
3	Pilling	LNFT	-
4	Guardrail	LNFT	23
5	Curb and Gutter	LNFT	-
6	Commercial Mix	TON	-
7	Plant Mix Surfacing	TON	1,787
8	Mobilization	LS	2
9	Milling and Pulverizing	SQYD	17,603
10	Crushed Aggregate Course	CUYD	-
11	Reinforcing Steel	LB	-
12	Excavation-unclassified	CUYD	-
13	Cover	SQYD	719,642
14	Riprap	CUYD	-
15	Traffic control	UNIT	458,931
16	Bridge Deck	CUYD	14,209
17	Signs	LS	1
18	Seeding	ACRE	-
19	Asphalt Cement	TON	113
20	Drainage Pipe (< 24 in)	LNFT	-
21	Special Borrow	CUYD	-
22	Sidewalk	SQYD	-
23	Drainage Pipe (> 24 in)	LNFT	-
24	Rumble Strips	ML	-
25	Emulsified Asphalt	GAL	649

The output section provides the project duration estimates calculated with two methods

Tool Maintenance and Database Update

- Data-driven prediction models may need to get updated over time and obtain new knowledge from recent projects to enhance prediction performance in the future
- The appendix explains steps and processes that need to be taken to update the model and the Excel tool using new data
- Processes to update the tool:



- For detailed explanation of updating the tool, refer to the final report

Thank you !

Questions and Comments