

# Machine Learning Workflow

Asking  
the right  
question

Preparing  
data

Selecting  
the  
algorithm

Training  
the  
model

Testing  
the  
model

# Machine Learning Workflow

Asking  
the right  
question

Preparing  
data

Selecting  
the  
algorithm

Training  
the  
model

Testing  
the  
model

Don't we  
already have  
the question?

“Predict if a person will develop diabetes”  
Need statement to direct and validate work  
Define end goal, starting point, and how to  
achieve goal

# Solution Statement Goals

Define scope (including data sources)

Define target performance

Define context for usage

Define how solution will be created



# Scope and Data Sources

*“Predict if a person will develop diabetes”*

*“Using Pima Indian Diabetes data, predict which people will develop diabetes”*

Understand the features in data

Identify critical features

Focus on at risk population

Select data source

Pima Indian Diabetes study is a good source



# Performance Targets

*“Using Pima Indian Diabetes data, predict which people will develop diabetes”*

*“Using Pima Indian Diabetes data, predict with 70% or great accuracy, which people will develop diabetes”*

**Binary result (True or False)**

**Coin Flip = 50% Accuracy**

**Genetic difference are a factor**

**70% Accuracy is common target**



# Context

*“Using Pima Indian Diabetes data, predict with 70% or great accuracy, which people will develop diabetes”*

*“Using Pima Indian Diabetes data, predict with 70% or great accuracy, which people are likely to develop diabetes.”*

Disease prediction

Medical research practices

Unknown variations between people

Likelihood is used



# Solution Creation

*“Using Pima Indian Diabetes data, predict with 70% or great accuracy, which people are likely to develop diabetes.”*

*“Use the Machine Learning Workflow to process and transform Pima Indian data to create a prediction model. This model must predict which people are likely to develop diabetes with 70% or greater accuracy.”*

## Machine Learning Workflow

- Process Pima Indian Data
- Transform data as required





*Use the Machine Learning Workflow to process and transform Pima Indian data to create a prediction model. This model must predict which people are likely to develop diabetes with 70% or greater accuracy.*



*Use the Machine Learning Workflow to process and transform **Pima Indian data** to create a prediction model. This model must predict which people are likely to develop diabetes with 70% or greater accuracy.*



*Use the Machine Learning Workflow to process and transform Pima Indian data to create a prediction model. This model must predict which people are likely to develop diabetes with 70% or greater accuracy.*



*Use the Machine Learning Workflow to process and transform Pima Indian data to create a prediction model. This model must predict which people are likely to develop diabetes with 70% or greater accuracy.*

