Workbook 2.1

Return to the insurance claims data based on the AnyCompany use case. Respond to the following questions:

1. Can you use a single AI service to address the problem?
   A. Yes
   B. No

Explain you're the reason for your answer below.

2. Can you use a combination of AI services?
   A. Yes
   B. No

Explain you're the reason for your answer below.

Course use case

- Global company offering many types of insurance
- Key challenge is fraud, which is costing millions of dollars per year
- Large amounts of historical data
- Want to improve new claims fraud prediction
- Resolution needed by end-of-year (EOY)

Meet Sofia Martinez. She's the Chief of Fraud Detection at AnyCompany. AnyCompany is an insurance company focused on many types of insurance. Recently, they have seen an increase in global insurance fraud that is costing them millions of dollars per year in financial losses, administrative overhead, and investigation activity. They have an enormous amount of historical data that could help them improve their process for predicting fraud for new claims and select the right rates for their customers. They are looking to work with you and your consulting company to address these challenges. They want to resolve this problem by the end of the calendar year with a new process.

As you progress through the course, you will learn how to apply machine learning to the fraud challenges faced by AnyCompany.

Workbook 3.1

While you complete this exercise, keep in mind that all machine learning problems start with data. Here is a snippet of data from AnyCompany showing car insurance claims. Remember, their goal is to predict whether a new insurance claim is fraudulent based on historical claims data, and they need a better process to determine claim validity. Look at the dataset on the next page, and answer the following questions.

1. Can you formulate the data and business problem as a machine learning problem?

   A. Yes
   B. No

2. Which type of machine learning can be used to address this problem?

   A. Regression
   B. Grouping/clustering
   C. Classification
   D. Reinforcement learning

3. Which columns would be considered features for the machine learning problem?

4. Which column would be considered the label for the machine learning problem?

AnyCompany Claims Data

	months_as_customer	age	policy_number	policy_bind_date	policy_state	policy_csl	policy_deductable	policy_annual_premium	umbrella_limit	insured_zip	police_report_available	total_claim_amount	injury_claim	property_claim	vehicle_claim	auto_make	auto_model	auto_year	fraud_reported
0	328	48	521585	2014-10-17	OH	250/500	1000	1406.91	0	466132	YES	71610	6510	13020	52080	Saab	92x	2004	Y
1	228	42	342868	2006-06-27	IN	250/500	2000	1197.22	5000000	468176	?	5070	780	780	3510	Mercedes	E400	2007	Y
2	134	29	687698	2000-09-06	OH	100/300	2000	1413.14	5000000	430632	NO	34650	7700	3850	23100	Dodge	RAM	2007	N
3	256	41	227811	1990-05-25	IL	250/500	2000	1415.74	6000000	608117	NO	63400	6340	6340	50720	Chevrolet	Tahoe	2014	Y
4	228	44	367455	2014-06-06	IL	500/1000	1000	1583.91	6000000	610706	NO	6500	1300	650	4550	Acura	RSX	2009	N

Workbook 4.1

Think about the insurance claims data based on the AnyCompany use case. The goal is to predict whether a new insurance claim is fraudulent based on historical claims data. AnyCompany needs a better process in the future to determine claim validity.

1. Which type of data is this – structured or unstructured?
   A. Yes
   B. No

2. Do all records have common fields?
   A. Yes
   B. No

3. Are all fields populated with data of the same type?
   A. Yes
   B. No

4. Is any data missing?
   A. Yes
   B. No

5. Which features would be most relevant for identifying fraudulent claims?

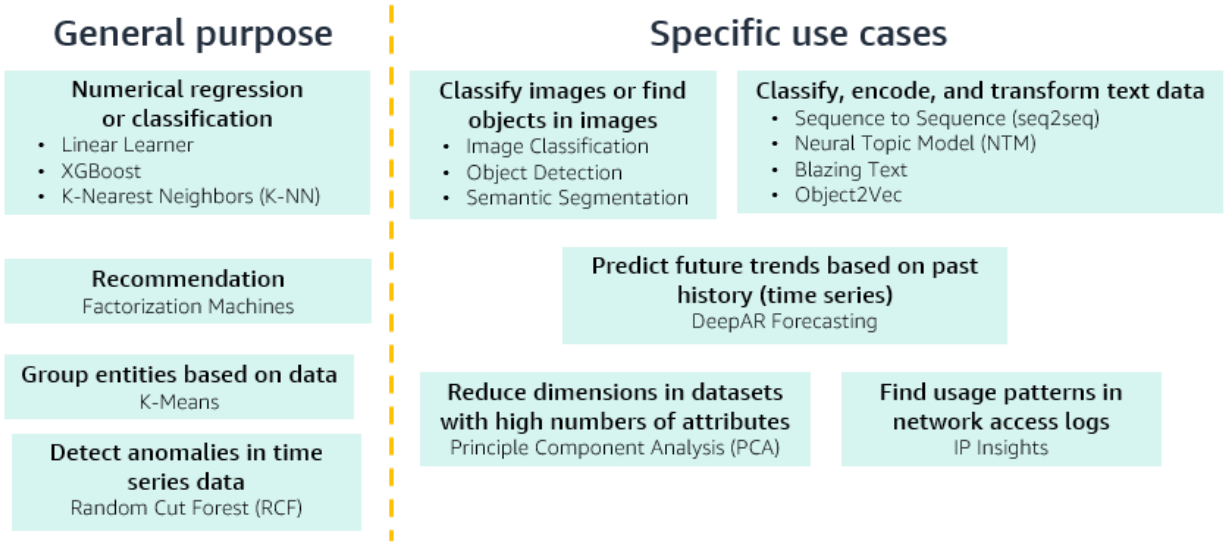
Workbook 5.1

Think about the AnyCompany Insurance Fraud use case.

Now that you have learned about AWS Deep Learning AMIs (DLAMIs), could a Deep Learning AMI be used to predict future insurance claim fraud? Provide information to support your answer.

Workbook 8.1

Return to the insurance claims data based on the AnyCompany use case. Refer to the Amazon SageMaker built-in algorithms shown here.



1. Which algorithms could be used to address the ML problem?

2. Why should AnyCompany use the selected algorithm?