2.7 Exercises

1. An online movie streaming company has a business problem of growing **customer churn**—subscription customers canceling their subscriptions to join a competitor. Create a list of ways in which predictive data analytics could be used to help address this business problem. For each proposed approach, describe the predictive model that will be built, how the model will be used by the business, and how using the model will help address the original business problem.

2. A national revenue commission performs audits on public companies to find and fine tax defaulters. To perform an audit, a tax inspector visits a company and spends a number of days scrutinizing the company's accounts. Because it takes so long and relies on experienced, expert tax inspectors, performing an audit is an expensive exercise. The revenue commission currently selects companies for audit at random. When an audit reveals that a company is complying with all tax requirements, there is a sense that the time spent performing the audit was wasted, and more important, that another business who is not tax compliant has been spared an investigation. The revenue commissioner would like to solve this problem by targeting audits at companies who are likely to be in breach of tax regulations, rather than selecting companies for audit at random. In this way the revenue commission hopes to maximize the yield from the audits that it performs.

To help with **situational fluency** for this scenario here is a brief outline of how companies interact with the revenue commission. When a company is formed, it registers with the company registrations office. Information provided at registration includes the type of industry the company is involved in, details of the directors of the company, and where the company is located. Once a company has been registered, it must provide a tax return at the end of every financial year. This includes all financial details of the company's operations during the year and is the basis of calculating the tax liability of a company. Public companies also must file public documents every year that outline how they have been performing, details of any changes in directorship, and so on.

a. Propose two ways in which predictive data analytics could be used to help address this business problem.[10] For each proposed approach, describe the predictive model that will be built, how the model will be used by the business, and how using the model will help address the original business problem.

b. For each analytics solution you have proposed for the revenue commission, outline the type of data that would be required.

c. For each analytics solution you have proposed, outline the capacity that the revenue commission would need in order to utilize the analytics-based insight that your solution would provide.

3. The table below shows a sample of a larger dataset containing details of policy holders at an insurance company. The descriptive features included in the table

describe each policy holders' ID, occupation, gender, age, the value of their car, the type of insurance policy they hold, and their preferred contact channel.

ID	OCCUPATION	GENDER	AGE	MOTOR VALUE	POLICY TYPE	PREF CHANNEL
1	lab tech	female	43	42,632	planC	sms
2	farmhand	female	57	22,096	planA	phone
3	biophysicist	male	21	27,221	planA	phone
4	sheriff	female	47	21,460	planB	phone
5	painter	male	55	13,976	planC	phone
6	manager	male	19	4,866	planA	email
7	geologist	male	51	12,759	planC	phone
8	messenger	male	49	15,672	planB	phone
9	nurse	female	18	16,399	planC	sms
10	fire inspector	male	47	14,767	planC	email

a. State whether each descriptive feature contains numeric, interval, ordinal, categorical, binary, or textual data.

b. How many levels does each categorical and ordinal feature have?

4. Select one of the predictive analytics models that you proposed in your answer to Question 2 about the revenue commission for exploration of the design of its **analytics base table** (**ABT**).

a. What is the prediction subject for the model that will be trained using this ABT?

b. Describe the domain concepts for this ABT.

c. Draw a domain concept diagram for the ABT.

d. Are there likely to be any legal issues associated with the domain concepts you have included?

✻ 5. Although their sales are reasonable, an online fashion retailer is struggling to generate the volume of sales that they had originally hoped for when launching their site. List a number of ways in which predictive data analytics could be used to help address this business problem. For each proposed approach, describe the predictive model that will be built, how the model will be used by the business, and how using the model will help address the original business problem.

✻ 6. An oil exploration company is struggling to cope with the number of exploratory sites that they need to drill in order to find locations for viable oil wells. There are many potential sites that geologists at the company have identified, but undertaking exploratory drilling at these sites is very expensive. If the company could increase the percentage of sites at which they perform exploratory drilling that actually lead to finding locations for viable wells, they could save a huge amount of money.

Currently geologists at the company identify potential drilling sites by manually examining information from a variety of different sources. These include ordinance survey maps, aerial photographs, characteristics of rock and soil samples taken from potential sites, and measurements from sensitive gravitational and seismic instruments.

a. Propose two ways in which predictive data analytics could be used to help address the problem that the oil exploration company is facing. For each proposed approach, describe the predictive model that will be built, how the model will be used by the company, and how using the model will help address the original problem.

b. For each analytics solution you have proposed, outline the type of data that would be required.

c. For each analytics solution you have proposed, outline the capacity that would be needed in order to utilize the analytics-based insight that your solution would provide.

✻ 7. Select one of the predictive analytics models that you proposed in your answer to the previous question about the oil exploration company for exploration of the design of its **analytics base table**.

a. What is the prediction subject for the model that will be trained using this ABT?

b. Describe the domain concepts for this ABT.

c. Draw a domain concept diagram for the ABT.

d. Are there likely to be any legal issues associated with the domain concepts you have included?

---

1 Remember that in insurance we don't refer to customers!

2 See the discussion in Section 2.1[21] relating to data availability, data connections, data granularity, data volume, and data time horizons.

3 It is important to remember for this discussion that all the data from which we construct an ABT for training and evaluating a model will be historical data.

4 Some might argue that the information on the application form summarizes an applicant's entire life, so this constitutes the observation period in this case!

5 The full text of the Civil Rights Act of 1964 is available at `www.gpo.gov/fdsys/granule/STATUTE-78/STATUTE-78-Pg241/content-detail.html`.