Academic English
Part II

Midterm Test
(sample)

MSU
Moscow
2012

Task 1: Put the following paragraphs in the right order to make a coherent text.

A

This result can be applied to a wide variety of data sets, including electricity bills[1], street addresses, lengths of rivers, physical and mathematical constants, and processes described by power laws. It tends to be very accurate when values are distributed across multiple orders of magnitude[2].

B

Benford's law can only be applied to data that is distributed across multiple orders of magnitude. For example, one might expect that Benford's law would apply to a list of numbers representing the populations of UK villages beginning with 'A'. But if a "village" is a settlement with population between 300 and 999, then Benford's law will not apply.

C

In 1972, it was suggested that the law could be used to detect possible fraud[3] in lists of socio-economic data submitted in support of public planning decisions. Based on the plausible[4] assumption that people who make up figures tend to distribute[5] their digits fairly uniformly, a simple comparison of first-digit frequency distribution from the data with the expected distribution according to Benford's law had to show up any anomalous results. Following this idea, Mark Nigrini showed that Benford's law could indicate accounting[6] and expenses fraud. In the United States, evidence[7] based on Benford's law is legally admissible[8] in criminal cases at all levels.

D

Benford's law, also called the **first-digit law**, states that in lists of numbers from many (but not all) real-life sources of data, the digit at the beginning is distributed in a specific, non-uniform way. According to it, the first digit is 1 about 30% of the time, and larger digits occur as the leading digit with lower frequency, to the point where 9 as a first digit occurs less than 5% of the time. This distribution of first digits is the same as the widths of gridlines[9] on the logarithmic scale.

---

[1] Счета.
[2] Большое количество величин порядков/величины нескольких порядков.
[3] Подделка, мошенничество.
[4] Правдоподобный, вероятный.
[5] Распределять.
[6] Бухгалтерский учет.
[7] Доказательство.
[8] Принимаемый в качестве доказательства (юр.).
[9] Линии (сетки).

Task 2: Shorten the text, crossing out unnecessary words/word-combinations/sentences so that the main idea remains the same (you can cross out 55 words maximum).

Machine learning is a branch of computer science that searches through data sets to make predictions about the future, ~~using the results of the search~~.

It is used to identify economic trends, personalize recommendations and ~~also to~~ build computers that appear to think ~~as well~~.

Although ~~however~~ machine learning has become incredibly popular, it only works on problems with large data sets, ~~and this technique cannot be applied to small data sets, as computers require a lot of information to build on~~.

~~Sometimes it might happen that the sequences and analogies found by the machine are in fact completely random, so the programmers should take such consequences into account.~~

Practitioners of machine learning must be careful while programming the machines so that they would not identify non-existing patterns.