

---

# A template for the arxiv style

---

A Preprint

Полина Юрьевна Кривуля  
Факультет ВМК  
МГУ имени М.В.Ломоносова  
Москва, Россия  
polina\_krivulya@mail.ru

Олег Валентинович Сенько  
Факультет ВМК  
МГУ имени М.В.Ломоносова  
Москва, Россия  
senkoov@mail.ru

## Abstract

В представленной работе осуществляется исследование задач, в которых целевая переменная определена через доверительные интервалы, а не посредством обучающей выборки. Основной целью данного исследования является изучение возможностей построения моделей для таких задач и проведение качественного анализа полученных результатов.

Одним из преимуществ данного способа определения целевой переменной является улучшение результатов по сравнению с моделями, использующими табличную обучающую выборку, благодаря возможности исключения выбросов. Также преимуществом данного подхода является его актуальность в случаях, когда наличие полноценной обучающей выборки невозможно.

В ходе исследования проводится математическое исследование различных методов построения моделей с использованием доверительных интервалов. Рассматриваются как классические статистические методы, так и алгоритмы машинного обучения. Данные методы применяются на эпидемиологических данных, проводится качественная оценка полученных результатов и сравнение их с моделями, обученными по классической табличной выборке.

Использование доверительных интервалов для определения целевой переменной способствует повышению надежности модели и уменьшению влияния случайных факторов на ее результаты. Данный подход особенно актуален в условиях, когда доступ к полноценным обучающим данным ограничен или невозможен.

Keywords Машинное обучение · Статистический анализ · Эпидемиология · Доверительные интервалы

## 1 Introduction

Острые респираторные вирусные инфекции (ОРВИ) — наиболее распространенная патология, на долю которой приходится около 90% всех инфекционных болезней. В Российской Федерации ежегодно болеют ОРВИ и гриппом более 30 млн человек, а суммарный экономический ущерб от респираторных вирусных инфекций оценивается в 40 млрд руб. в год. Большие экономические потери обусловлены вовлечением в эпидемический процесс трудоспособного населения, развитием осложнений, непродолжительным нестойким иммунитетом, определяющим повторные случаи заболевания.

Вирусы, вызывающие респираторные инфекции, не являются эндемичными для какого-либо региона или страны и распространены повсеместно. Чаше эпидемии возникают в зимнее время, однако вспышки наблюдаются и в осенне-весенний период, а спорадические случаи ОРВИ — круглый год. Известно около 300 возбудителей респираторных инфекций, более 200 из них представители 4 семейств РНК-содержащих вирусов (ортомиксовирусы, парамиксовирусы, коронавирусы и пикорнавирусы) и 2 семейств ДНК-содержащих вирусов (аденовирусы и герпесвирусы).

Цель данного исследования заключается в анализе и оценке характеристик нескольких типов вирусов, а именно гриппа, аденовируса, респираторно-синцитиального вируса (РС) и бокавируса. Для достижения данной цели проведен анализ набора симптомов, характерных для каждого из этих вирусов, а также учтены сезонные особенности. Ключевыми этапами данного исследования являются оценка статистической значимости симптомов и вычисление априорных и апостериорных вероятностей наличия каждого типа вируса. Для дальнейшего распознавания типа вируса по набору признаков использован алгоритм наивного байесовского классификатора. Результаты данного исследования могут быть полезны для лучшего понимания характеристик и распространения данных вирусов, а также для разработки эффективных методов профилактики и лечения.

## 2 Задача

В данной работе использованы четыре таблицы с информацией о характеристиках гриппа, аденовируса, респираторно-синцитиального вируса и бокавируса. Симптомы, рассматриваемые в данном исследовании, включают температуру, кашель, одышку, боли в горле, боль в груди, ринит, головную боль, диарею и боли в животе. Информация о суммарном числе заболевших и количестве людей, имеющих каждый из симптомов, была собрана из статей. Важно отметить, что в некоторых статьях отсутствует информация о наличии части симптомов у пациентов. В частности, отсутствует информация о симптоме "Боль в груди" у аденовируса во всех статьях, а также "Боль в груди" и "Головная боль" у бокавируса.

Для получения априорной вероятности наличия каждого типа вируса по сезонам использована отдельная таблица, отражающая заболеваемость каждым из вирусов по датам.

Заметим, что особенность данных заключается в отсутствии обучающей выборки.

### 2.1 Headings: second level

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

$$\xi_{ij}(t) = P(x_t = i, x_{t+1} = j | y, v, w; \theta) = \frac{\alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}{\sum_{i=1}^N \sum_{j=1}^N \alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})} \quad (1)$$

#### 2.1.1 Headings: third level

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Paragraph Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

## 3 Examples of citations, figures, tables, references

### 3.1 Citations

Citations use **natbib**. The documentation may be found at



Рис. 1: Sample figure caption.

<http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf>

Here is an example usage of the two main commands (`citet` and `citep`): Some people thought a thing [Kour and Saabne, 2014a, Hadash et al., 2018] but other people thought something else [Kour and Saabne, 2014b]. Many people have speculated that if we knew exactly why Kour and Saabne [2014b] thought this...

### 3.2 Figures

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi. See Figure 1. Here is how you add footnotes.<sup>1</sup> Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

### 3.3 Tables

See awesome Table 1.

The documentation for `booktabs` ('Publication quality tables in LaTeX') is available from:

<https://www.ctan.org/pkg/booktabs>

### 3.4 Lists

- Lorem ipsum dolor sit amet
- consectetur adipiscing elit.
- Aliquam dignissim blandit est, in dictum tortor gravida eget. In ac rutrum magna.

---

<sup>1</sup>Sample of the first footnote.

Таблица 1: Sample table title

Part		
Name	Description	Size ( $\mu\text{m}$ )
Dendrite	Input terminal	$\sim 100$
Axon	Output terminal	$\sim 10$
Soma	Cell body	up to $10^6$

### Список литературы

- George Kour and Raid Saabne. Real-time segmentation of on-line handwritten arabic script. In *Frontiers in Handwriting Recognition (ICFHR)*, 2014 14th International Conference on, pages 417–422. IEEE, 2014a.
- Guy Hadash, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, and Alon Jacovi. Estimate and replace: A novel approach to integrating deep neural networks with existing applications. *arXiv preprint arXiv:1804.09028*, 2018.
- George Kour and Raid Saabne. Fast classification of handwritten on-line arabic characters. In *Soft Computing and Pattern Recognition (SoCPaR)*, 2014 6th International Conference of, pages 312–318. IEEE, 2014b. doi:10.1109/SOCPAR.2014.7008025.