**Вакансия: Научный сотрудник PostDoc**

**г. Пермь**

**Название проекта:** Выделение подгрупп специфичных наблюдений данных

**Подразделение:** Научно-учебная лаборатория междисциплинарных эмпирических исследований

Мы ищем энергичного, нацеленного на эффективную научную работу коллегу, интересующегося анализом данных и машинным обучением, желающего разрабатывать новые интерпретируемые методы машинного обучения.

Интерпретируемость методов машинного обучения необходима для обнаружения новых феноменов и лучшего понимания процесса, описываемого выборкой данных. Один из интерпретируемых подходов в машинном обучении представлен методами класса Subgroup Discovery, которые из всей выборки данных на основании независимых переменных выделяют подгруппы наблюдений, имеющих существенно отличные от других наблюдений характеристики. Такую задачу можно формулировать как в условиях обучения без учителя, и тогда необходимо обнаружить подгруппу наблюдений, распределение которых существенно отличается от некоторого ожидаемого. В этом случае задача становится близкой задаче кластеризации, но при этом не зависит от функции расстояния между наблюдениями. Также эту задачу можно формулировать в терминах обучения с учителем. В этом случае обнаруживаемая подгруппа должна существенно отличаться от остальных наблюдений с точки зрения целевой переменной.

На практике такие методы могут использоваться, например, в задачах персонализированного маркетинга, отвечая на вопрос: для каких клиентов нужно разрабатывать новую маркетинговую кампанию.

Цель проекта: Разработка, сравнение и апробация на реальных данных новых методов обнаружения специфичных подгрупп наблюдений для разных контекстов.

Задачи в рамках проекта:

* Разработка и реализация метода класса Subgroup discovery, для выявления подгрупп наблюдений выборки данных, подчиняющихся распределению отличному от усредненного.
* Проверка качества разработанного метода на синтетических и реальных данных.
* Сравнение разработанного метода с другими интерпретируемыми методами машинного обучения, включая деревья решений и методы кластеризации.

Заработная плата: 70 000 до вычета НДФЛ

Место работы: основное

Ставка: 1

Срок работы: 1 год с возможностью продления

### Условия:

* работа в заинтересованном молодом коллективе
* участия в программах академической мобильности и программах повышения квалификации
* участие в научных и образовательных мероприятиях и программах НИУ ВШЭ
* работа на реальных данных, представленных ведущими российскими и международными компаниями

### Требования:

* российское гражданство
* возраст не более 35 лет
* отсутствие трудовых или учебных отношений с НИУ ВШЭ
* ученая степень кандидата наук (или успешная защита кандидатской диссертации) или степень PhD, полученная в российском или иностранном университете.
* высокая мотивация, желание осваивать новое;
* знание английского языка;
* знание статистики;
* владение статистическими методами анализа данных;
* владения языком программирования, предпочтительно R, Python или C/C++.