

# elect 式 value-tracker

e.tmailbank@gmail.com

## 1 最終目標

公開できる〇〇予測 API にする。

## 2 担当教員に求めたいもの

大規模データの機械学習、データベース処理に関する知識。  
また、それらをつなげるためのメソッド。  
疑問点ができ次第 質問させていただきたいと思っています。

### 2.1 現在質問したい内容

- 大規模なデータを毎秒単位で収集しなければならない可能性がでてくるが、どうやってコンピュータで処理すべきなのか（メモリ等の問題）
- 大容量のデータを処理できるデータベースにはどのようなものがふさわしいか（Hbase でこと足りるのか）
- データベースと機械学習をつなげるパイプをどのように設計すればいいのか
- 機械学習全般の知識 (学習は主には自分たちでついていく予定です)

## 3 担当教員に求めないもの

プログラミング言語の教育技術（使う言語は自分達で学習）  
全体のスケジュール管理（グループリーダーが統括する予定です）

## 4 目標を絞った現段階での目標

すべての一>何か に

- 例えば...  
コンビニの売上予想 … 必要なデータ例

内容	気温	湿度	仕入れ	交通量	株価	人口遷移
重み	5	4	2	5.5	0.1	0.2

### 4.1 データの入手

はじめのデータは既存のデータ提供サービスから入手する。例えば、json などの形式で入手した後、それをデータベースに登録する。最終的には利用者から提供されたデータ（試行に用いられた）データなどの利用も視野に入りたい。

### 4.2 検索

関連性のあるデータを計算に入れることにするが、その関連性のあるデータの選択については現在検討中である。（利用者に選択してもらうスタイルを取る、何らかの学習でサジェストするなど）

### 4.3 データの出力

データの出力は利用者の入れたデータに関連して、利用者の求めたデータを返す。

## 5 具体的な内容を詰めてみる。

### 5.1 データベース

#### 5.1.1 データベースにほしいもの

より高精度な計算のためあらゆるデータが欲しい（※すべてのデータを計算するかどうかは別）

#### 5.1.2 データをどうやって集めるのか

前項で述べたとおり、初期段階では例えば weather.api などからの公開されているデータを、それから各利用者から与えられるデータを集める。

### **5.1.3 データベースを管理するに当たって**

Hbase などの導入を考えている。

### **5.1.4 処理するデータの量**

取捨選択して毎 s～毎日に合計で GB 単位で流し込んでみたい。

## **5.2 機械学習**

### **5.2.1 機械学習とは**

初学者ながら、TensorFlow という機械学習のツールを用いることを考えている。他に  
ある素晴らしいツールが目によれば、それを用いる。

### **5.2.2 処理するデータの大きさは**

GB 単位になると予想している。

### **5.2.3 与えられるデータは**

主に数値データを扱う。