

## 進捗報告

]

### 1 今週やったこと

- 交叉検定
- 不均衡データの処理

### 2 交叉検定

テンパイ予測モデルで 4 分交叉検定をした。表 1 に結果を示す。

表 1: 交叉検定

	1	2	3	4	平均
0(ノーテン) の F1	0.9576	0.9591	0.9576	0.9570	0.957825
1(テンパイ) の F1	0.5693	0.5767	0.5670	0.5725	0.571375
acc	0.9229	0.9254	0.9228	0.9218	0.923225

### 3 不均衡データの処理

不均衡データの処理は二つある、重みつけとサンプリング。元データで学習する時は予測率高くても再現率が低い結果。アップサンプリングと重みつけると予測率が低くても再現率が高い結果になる。予測率と再現率のバランスを取るために、データをアップサンプリングしたデータ (1:3.34) を使って、テンパイに小さい重み (1:0.75) をつけて、バランスを取った結果を表 2 に示す。

表 2: 不均衡データ処理後の結果

	precision	recall	f1	support
0(ノーテン)	0.9587	0.9576	0.9581	62835
1(テンパイ)	0.6084	0.6151	0.6117	6735
acc			0.9244	69570