**Project Proposal：Personal Income Analysis and Prediction**

**ECE180 Team2**

Kaiyu Zhang

**PURPOSE**

This project aims to predict the possibility whether a person’s income is above 50K or not, which will help employees consider the suitable salary for the new worker, and also help people evaluate their income level in a more considerable way. According to Trading Economics reports, annual salary above 50K will be define as High pay. Thus personal annual salary is above 50K or not is an important line to measure their living standards. We obtain the data from UC Irvine machine learning data-bases, which including the samples’ different features and their annual income is above 50K or not. Our analysis based on people’s age, company type, marriage, ethnic, gender, education, the number of family members, etc.

**DATA ACQUISITION AND ANALYSIS**

We obtain the data from UC Irvine machine learning data-bases, which including the samples’ different features and their annual income is above 50K or not. For processing these data, we use Python build-in libraries such as: Numpy, Pandas, Matplotlib, basemap and Scikit-learn. Matplotlib can help with the visualization. Basemap deals with associating geometric data with flights. Numpy and Pandas help to organize data and extract the statistics information of the data in a fast fashion. Scikit-learn will provide machine learning algorithms that we might use. By these libraries we can find the weights of personal features, and how it will affect with their income.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Time line** | **Kaiyu**  **Zhang** | **Wenyu**  **Zhang** | **Linqi Pan** | **Fengcan Zhu** | **Yucheng Huang** | **Woen Lee** | **Andrew(Dung) Luo** |
| **02/09-02/16** | Download Data | Organize Data | **Organize Data** |  |  |  |  |
| **02/17-02/24** | Exact Features |  |  |  |  |  |  |
| **02/25-03/03** | Make Visual Graphs |  |  |  |  |  |  |
| **03/04-03/11** | Prepare for Presentation |  |  |  |  |  |  |