**Nitrogen Concentration Prediction in Waste Water Study**

**(Math 571 Project Final Report)**

**Submitted to:**

Prof. Adam McElhinney Department of Data Science Illinois Institute of Technology

**Report Prepared By:**

Travis Boltz

A20389533

Yitao Ma

A20380311

Yuqing Zhao

A20392575

Yue Ning

A20282614

April 15, 201

**Table of Content**

## Executive summary ………………………………………………… i

## 1 Introduction ………………………………………………………. 4

## 2 Data Processing…...………………………………………………. xx

## Basic summary statistics ………………………………………….. xx

## Analysis of Demographic and Avova Test..………………………. xx

## Exploratory Regression Models ………………………………….. xx

## Evaluations of Models…………………………… ………………. xx

## Results and Discussion……………………...……………………… xx

## 8 References ……………………………………………………….. xx

## Appendix: R code…………….………………………………………. xx

**Executive Summary**

This report summarizes the statistical modeling and analysis results associated with the Nitrogen concentration prediction in waste water study. The purpose of this report is to document both the implemented sampling design and all corresponding data modeling and inference techniques used during the subsequent statistical analyses.

The development of the sampling protocol, including both the data source and data cleaning strategy are discussed in Section 2.

The basic statistics that summarize the contamination data associated with the analyzed compounds are given in Section 3. A total of 9 variables were analyzed for eight locations for this study. Five compounds concentration measurements include BOD5, TKN, NH3.N, P.TOT, SS. Two variables factors analyzed in this study include FLOW and Rainfall. The statistics summary and several plots were performed in this study.

Section 4 presents the analysis of the demographic, based on the 8 waste water treatment plants where data were acquired. ANOVA-test p-value approach were used to compare data for TKN (target variable) to see if the means are equal.

Regression models and Algorithm background are discussed in Section 5; both the baseline-linear regression and PCA may also be present; Additionally, Linear regression with variable selection and linear regression with regularization were used to XXXX.

Section 6 introduces model evaluations.

Finally, in section 7, discussion and conclusion are given.

**1.0 Introduction**

Sewage is an essential consideration for every part of the world. It is a water-carried waste, the water leftover after its use in numerous application such as industrial, agricultural, municipal, domestic and on. It had become a significant problem for the environment, especially when untreated sewage discharged into the body of water such as stream, river, lake, bay or ocean [1]. The ecosystem cannot breakdown such massive amount of human waste; sewage can contaminate water and harm vast amounts of wildlife. Sewage with no treatment unbalances the natural nutrient cycle by loading large concentrations of a complex mixture of chemicals with many distinctive chemical characteristics, like ammonium, nitrate, nitrogen, phosphorus and high dissolved solids into the water. Particularly phosphorus and nitrogen cause increased the growth of algae and green plants in the water. As more algae and plants grow, others die. The dead organic matter becomes bacteria’s food, with the increasing of food, the bacteria increase in number and use up the dissolved oxygen in the water, which results in a dead area [2]. A harmful algal blooms (HABs) is dangerous for the environment and human health and usually necessitates the shut-down of water treatment plants used for drinking water, which has massive consequences for the local economy and the sustainability of natural ecosystems. One of the ways to prevent these algae blooms is to understand better how much nitrogen is present in the natural ecosystem. The metric used to measure the total amount of nitrogen in the water is called Total Kjeldahl Nitrogen (TKN). TKN is the U.S. EPA-approved parameter used to measure organic nitrogen and ammonia. The TKN content of influent municipal wastewater is typically between 35 and 60 mg/L. Organic nitrogen compounds in wastewater undergo microbial conversion to NH3 and ammonium ion NH4+ [3]. This process requires a lab to analyze a sample of water to determine the TKN presents; only approximate calculation was provided. It is not always the case that TKN was captured in present or past data. However, TKN is a required parameter for regulatory reporting at many wastewater treatment plants for monitoring plants operations. To overcome this issue, we want to predict the level of TKN present in the water using other measurements that are present in the dataset.

**2.0 Data Processing**

**3.0 Basic Summary Statistics**

**4.0 Analysis of Demographic and Avova Test**

# 5.0 Exploratory Regression Models

**6.0 Evaluations of Models**

**7.0 Results and Discussion**

**8.0 References**

**Appendix: R Code**

(for performing all data analyses described in Report)