Universit´e Internationale de Rabat

E´cole Supérieur d’informatique et du Numérique

Cycle Ingénieur en Génie Informatique



**Machine Learning Project Report**

Fulfillment of the requirement of the degree of

#### Computer Science Engineering Diploma

**Theme:**

Richter’s Predictor Modeling Earthquake Damage

**Author:**

Bouziane Yassine

Tooogrra Yassine

Falhi Mohammed Reda

**Pedagogic advisors :**

Nada Sbihi, RIM-CRISTAL Lab., ENSI (Manouba, Tunisia)

[abdelfattah.belghith@ensi.rnu.tn](mailto:abdelfattah.belghith@ensi.rnu.tn)

Ghogho lbawal, RSM Dept., ENST-Bretagne (Rennes, France)

[jm.bonnin@enst-bretagne.fr](mailto:jm.bonnin@enst-bretagne.fr)

**Presented on june, 2022**

#### Abstract

#### R´esum´e

## Dedication

To my parents: words are not expressive To Salam : my sister and best friend

To MedAli et Touha: hoping that it will be a source of inspiration for them

End of Studies Project Report

October 2006,E´cole Nationale des Sciences de l’Informatique

## Greeting

First, I want to thank Dr. Nicolas Montavont for his faith in me and his sug- gestions. He was really a continual source of learning during all our discussion and meeting.

I would like to express my deepest gratitude to Mohamed Kassab, my friend and supervisor for his good suggestions, his support, and his good mood.

I really appreciated the critical comments and beneficial suggestions from my great Professor Abdelfattah Belghith.

A particular acknowledgement is for all my professors and teachers during these latter three years during my studies within the ENSI. They were an inex- haustible source of learning.

I would like to thank Amine Dhraief for his precious help at the starting of this project, and particularly in NS-2 discovery and study.

I would like to thank all those who works in the ENST-Bretagne for the cordial reception they reserved to us.

All my thankfulness is for my friends at Rennes. They provided me with a great comfort with their kindness, favour, generosity and humour. My thoughts go to Neila, Ray`ene, Meriem, Anis and all others.

For all my good friends in the ENSI: Sofien, Hamdi, Med, Youssef, Dina, Zied, Fadi, Walid, Ramsis and all others: I’ll never forget you.

My thoughts go to all those who have been a real help for me, my best friends: Emeni, Baha, Sana, Amine, Faten, . . .

Finally, my special thank goes to Attoufa and Mabrouk, for their continuous support, to aunt Fawzia and uncle Rachid, and to all my big family’s members. I love you.

Safa`a Hachana

# Contents

|  |  |  |
| --- | --- | --- |
|  | **Introduction** | **8** |
| **I** | **Data analysis** | **11** |
| **1** | **Preparing Data** | **12** |

* 1. Introduction 12
  2. Importing Data 12
  3. Data Exploration 13
     1. Check for missing values…………………………………………………..….. 14
     2. Check Correlation among features……………………………………….. 14
     3. Check distribution of interval features………………………………….. 14
     4. Analyzing features ………………………………………………………………. 14
     5. Test statistics………………………………………………………………………. 14
  4. Conclusion 15

|  |  |  |
| --- | --- | --- |
| **2** | **Data Modelling** | **12** |

* 1. Logistic Regression 12
  2. Random Forest 12
  3. Xgboost 12
  4. Gradient Boosting 12
  5. Conclusion 15

|  |  |
| --- | --- |
| **II Experiments and Results** | **11** |

|  |  |  |
| --- | --- | --- |
| **1** | **Preparing Data** | **12** |

* 1. Introduction 12
  2. Importing Data 12
  3. Data Exploration 13
     1. Check for missing values…………………………………………………..….. 14
     2. Check Correlation among features……………………………………….. 14
     3. Check distribution of interval features………………………………….. 14
     4. Analyzing features ………………………………………………………………. 14
     5. Test statistics………………………………………………………………………. 14
  4. Conclusion 15

**List of Figures**

* 1. Scheduler by NS-2 21

## Introduction

This project is about to work on the competition hosted by drivendata.com with the name “Ritcher’s Predictor: Modeling Earthquake Damage”

Based on aspects of building location and construction, the goal is to predict the level of damage to buildings caused by the 2015 Gorkha earthquake in Nepal.

The Central Bureau of Statistics that work under the National Planning Commission Secretariat of Nepal collected the data through surveys. This survey is one of the largest post-disaster datasets ever collected, containing valuable information on earthquake impacts, household conditions, and socio-economic-demographic statistics.

We're trying to predict the ordinal variable damage\_grade, which represents a level of damage to the building that was hit by the earthquake. There are 3 grades of the damage:

1. represents low damage
2. represents a medium amount of damage
3. represents almost complete destruction

**Part I**

# Data Analysis

**Chapter 1**

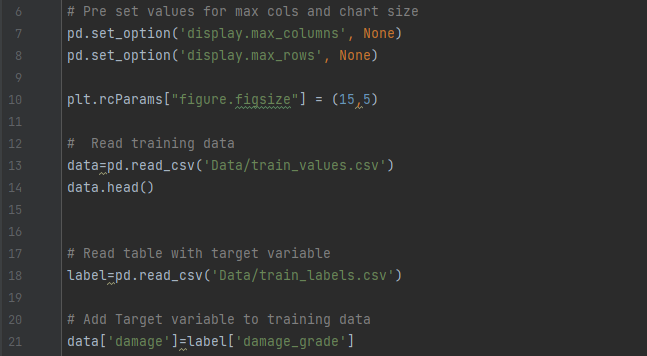
**Preparing Data**

## Introduction

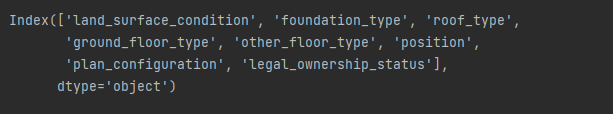
This section aims at preparing data after collecting it from a reliable source. Good data is relevant, contains very few missing and repeated values, and has a good representation of the various subcategories/classes present. This introductive part begins with importing data, exploring its structural components, features, and generate graphs.

## Importing and Preprocessing Data

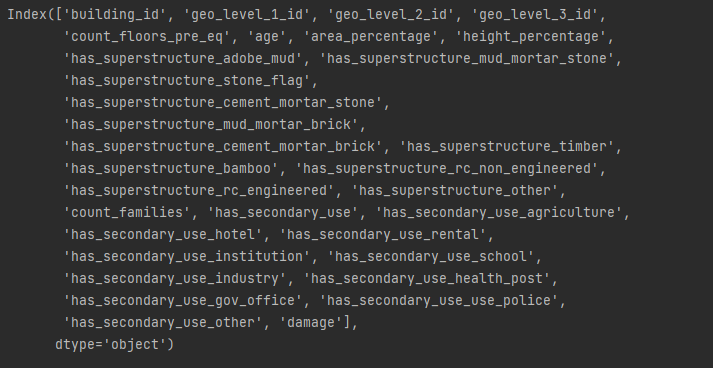
The dataset consists of information about household conditions and socio-economic-demographic statistics of the buildings damaged during Nepal’s Gorkha earthquake in 2015



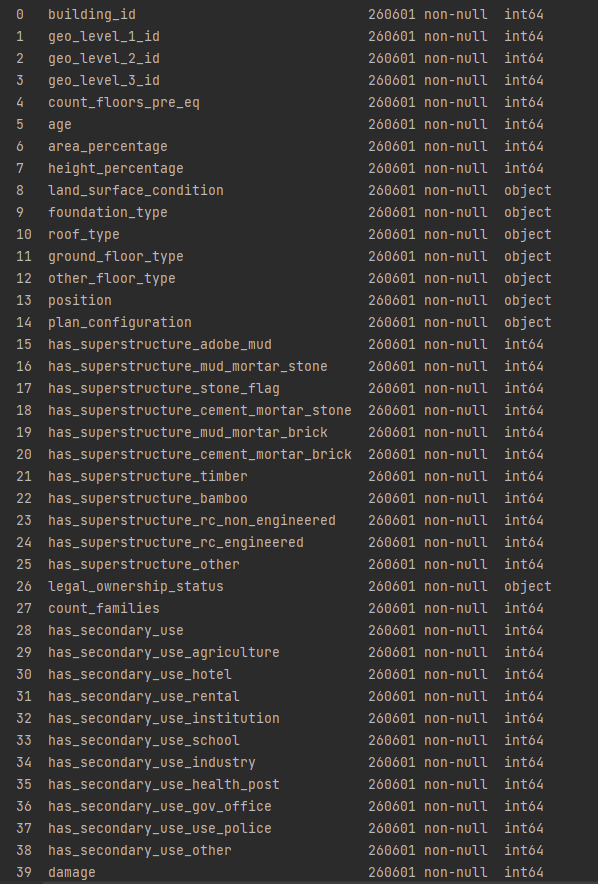
* The dataset mainly consists of information on the buildings' structure and their legal ownership. Each row in the dataset represents a specific building in the region that was hit by Gorkha earthquake.
* There are 39 columns in this dataset, where the **building\_id** column is a unique and random identifier. The remaining 38 features are described in the section below. **Categorical variables** have been obfuscated random lowercase ascii characters. The appearance of the same character in distinct columns does not imply the same original value.
* **Categorical Variables :**



* **Numeric Variables :**



* **Checking for missing values:** ALL the values are filled in, no missing values

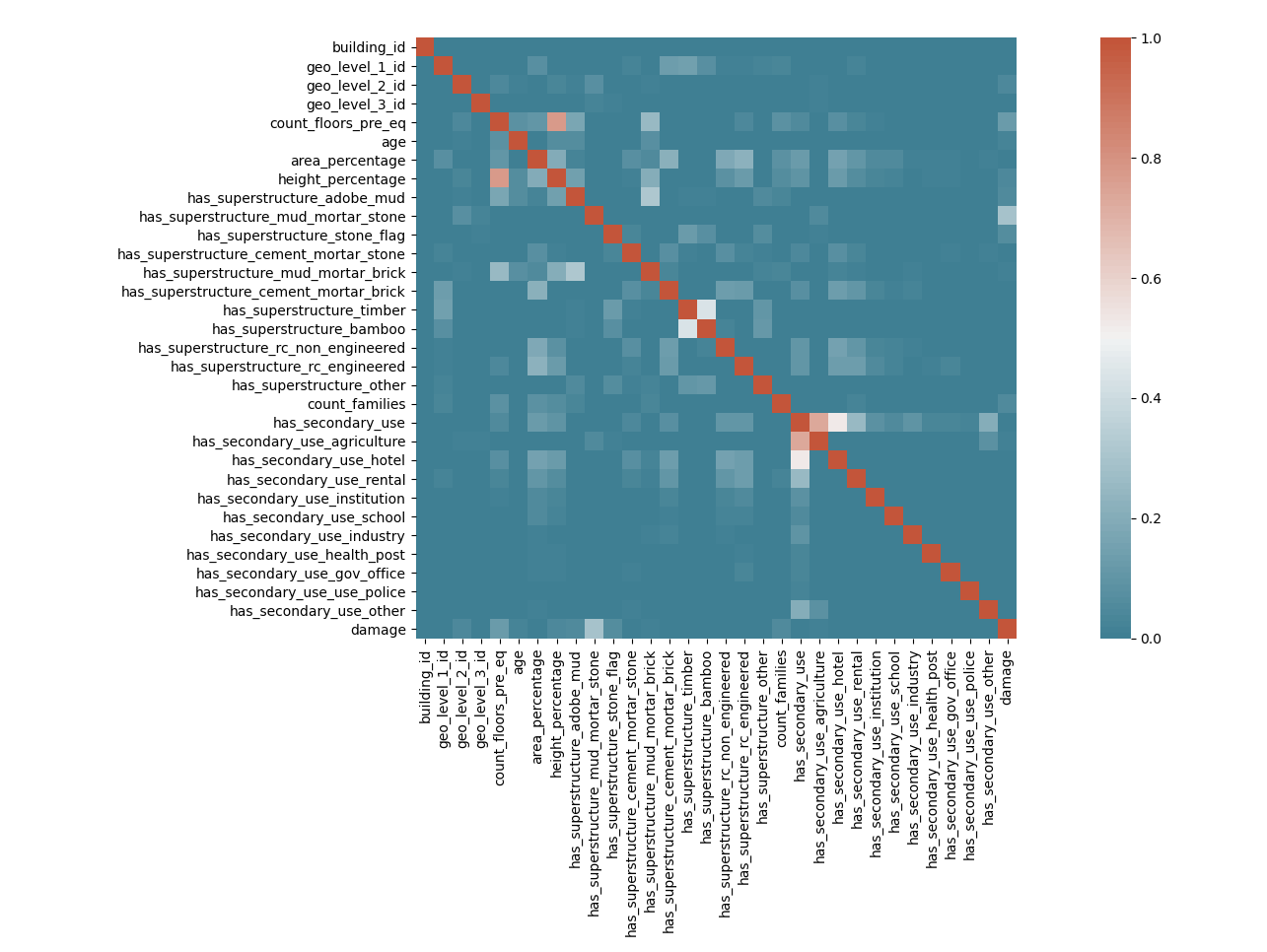


## Data Exploration

### Checking for correlation among features

### 

* Plotting the result :



* Features Correlation Result from plot:

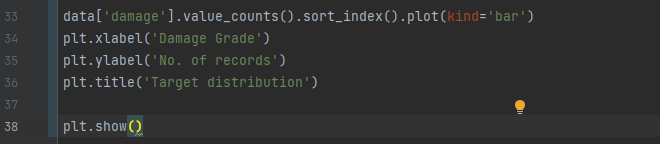
**Has\_secondary\_use** && **it’s sub\_parts**

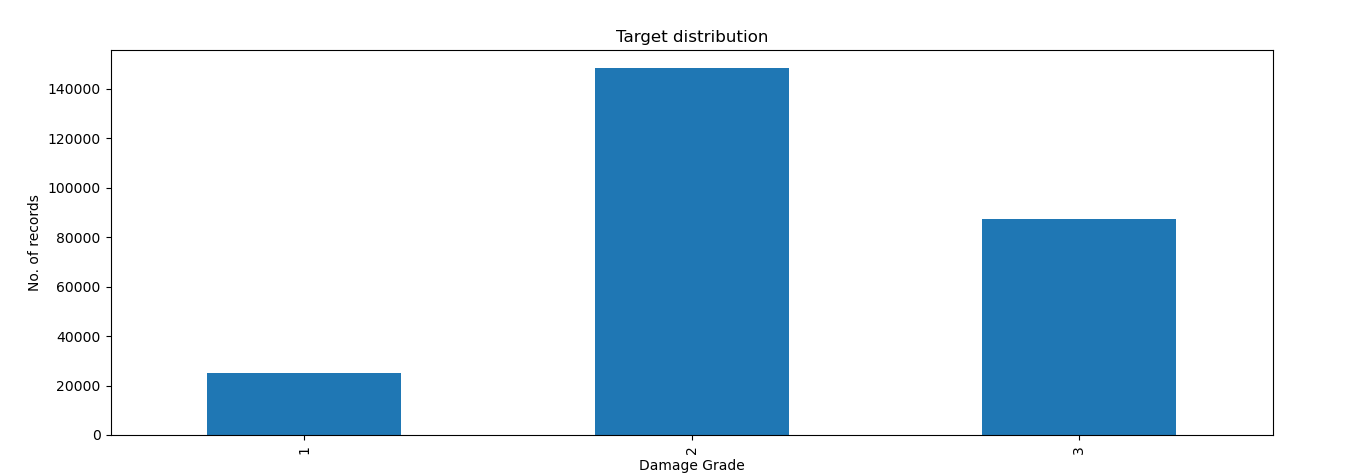
**Height\_percentage** && **count\_floors\_pre\_eq**

**Area\_percentage - height\_percentage** && **has\_super\_structure and seconday use of buildings**

### Checking for distribution of interval features

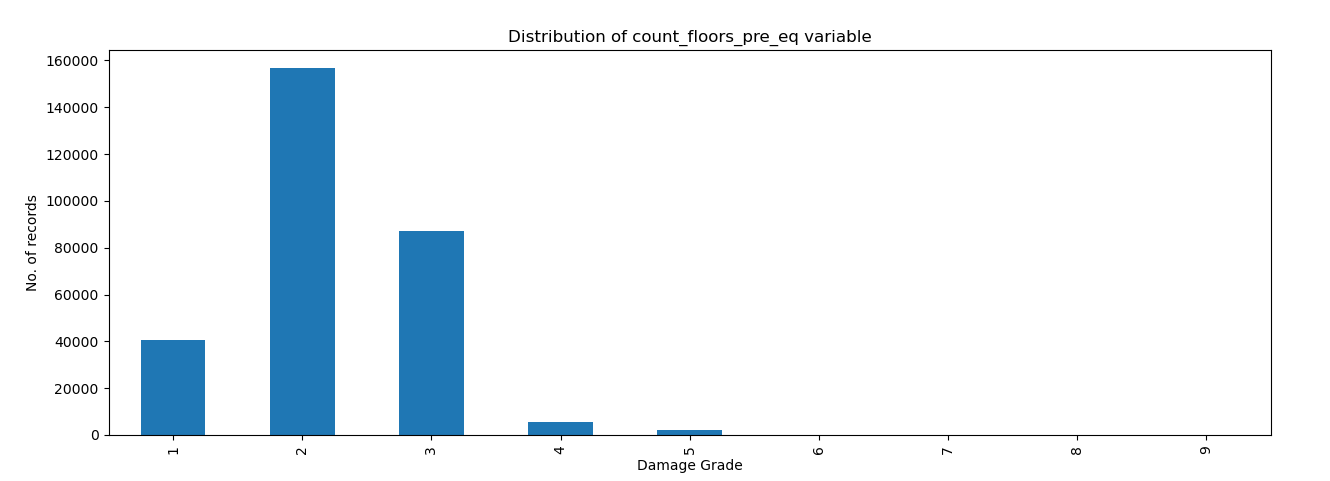
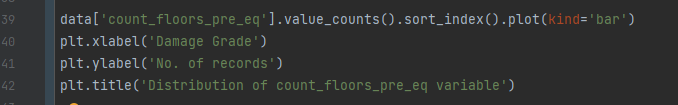
### Damage\_grade





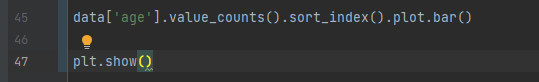
* Most of data points are towards **damage\_grade 2,** which can cause bais

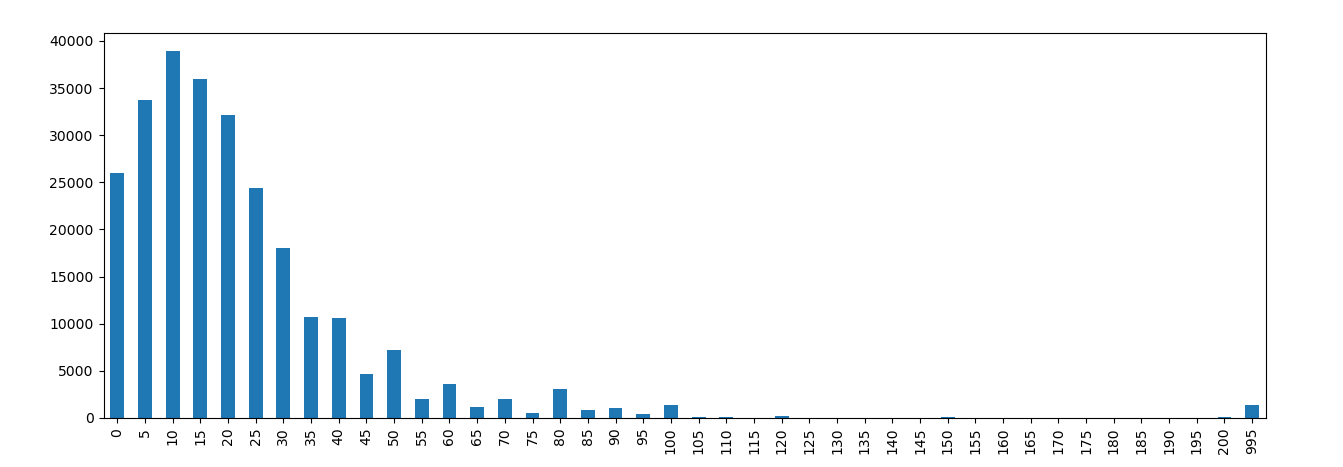
### Count\_floors\_pre\_eq



* Maximum damage happened to **2-storey** buildings

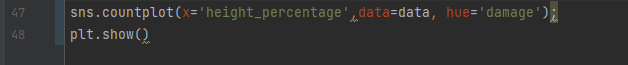
### Age

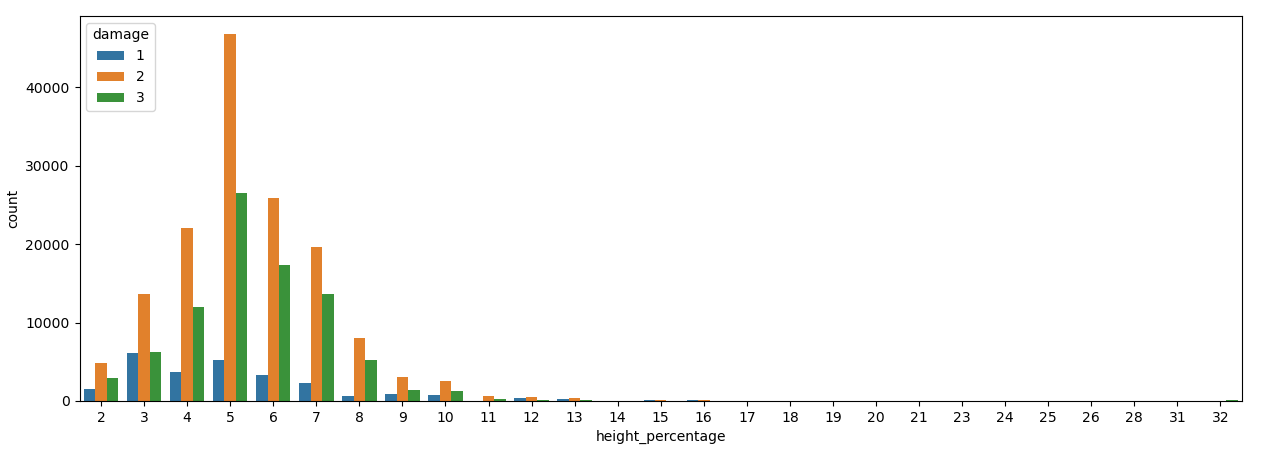




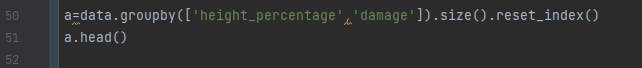
* Around 10% of total data has buildings with **age 0 to 4,** the values after the 100 years of age are excluded from the analysis

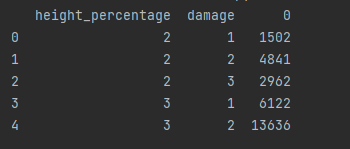
### Area\_percentage

**

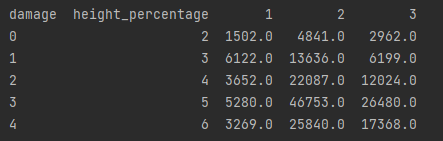


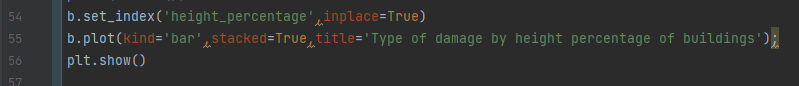
### Height\_percentage

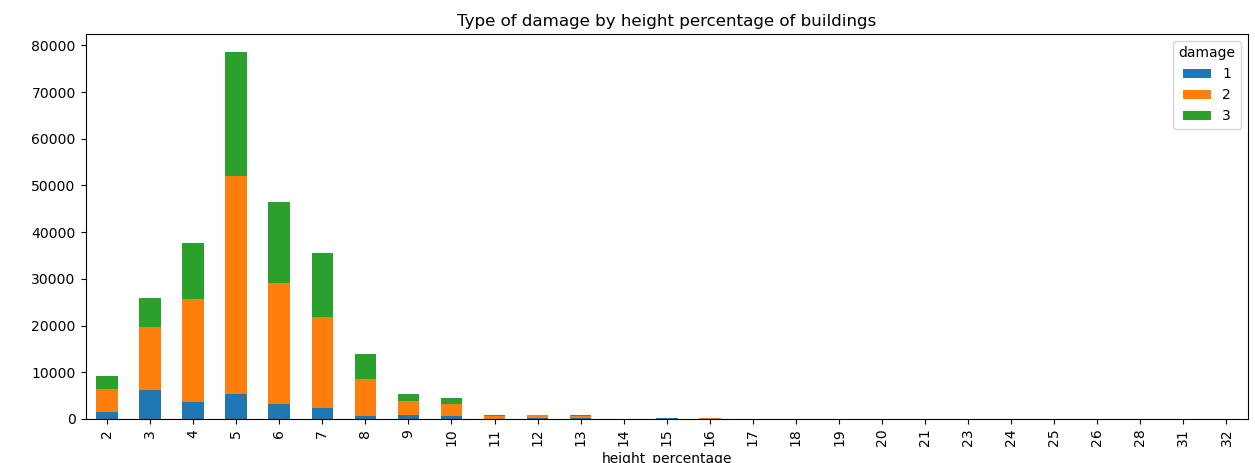
**



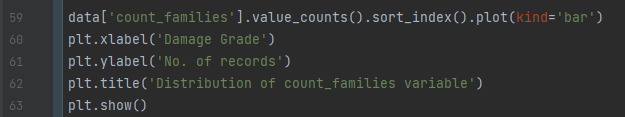


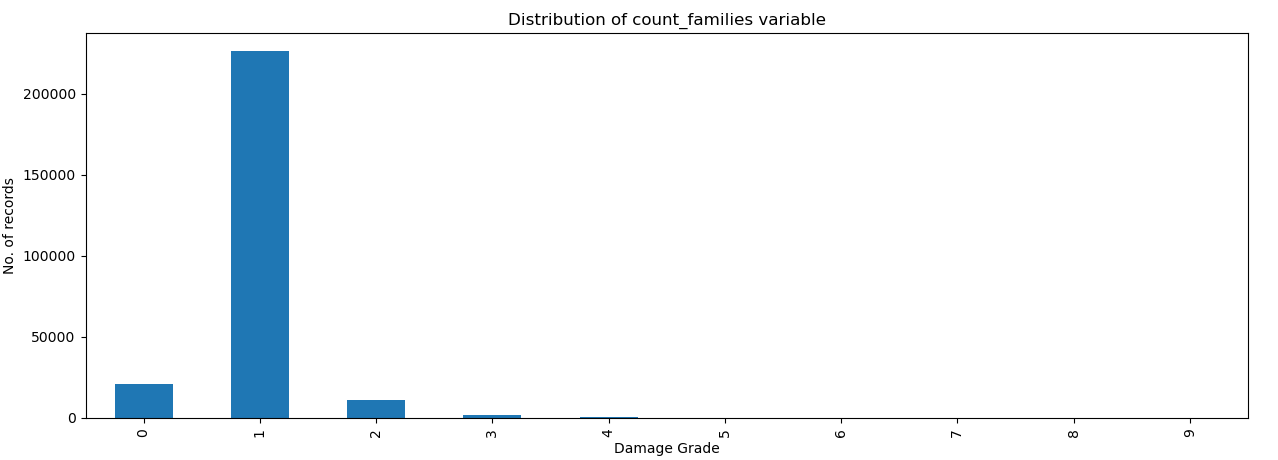






### Count\_families



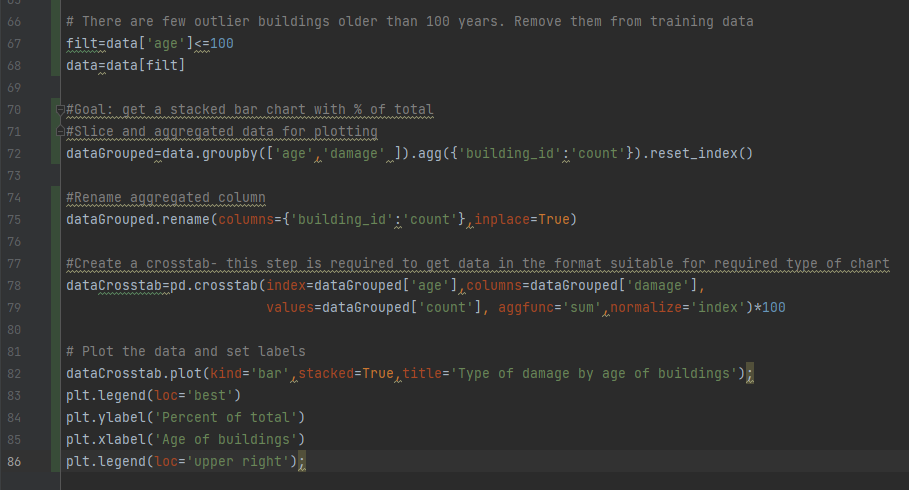


* Maximum homes damaged were **single faily homes**

### Analyzing features

### Damage\_grade

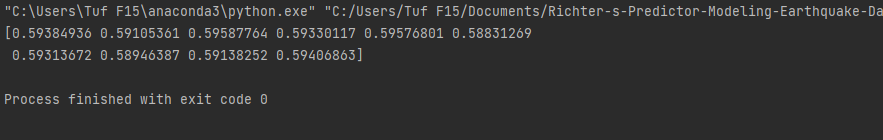
### 



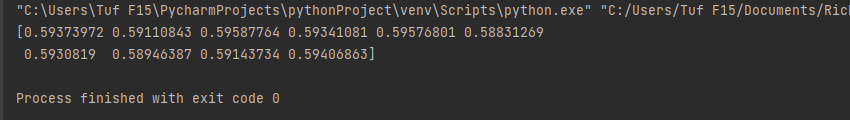
**Part II**

# Data Modeling

**After using pipeline**



**After using pipeline**

****

**Chapter 8**

# Conclusion and prospects

Along this report, we presented the different steps of development of a new wireless networks simulation environment.

This project begun with an initiation to event discrete simulation, giving a brief terminology that provides the reader with basic technical background. The sate of the art includes also an overview of different commonly used simulation tools. For this purpose, we made a comparative study of three simulators, namely NS-2, OMNeT++ and SimulX. Of course, we gave first the evaluation criteria on which we based in the comparison.

Starting from this evaluation, we analysed and detailed the issues raising from the use of these available simulators. Consequently, we notified that net- working research community needs a new simulation tool, that would be special- ized in wireless networks and mobility features, and will provide some specific functionalities and properties such as extensibility, total modularity, scalability, easiness of use for pedagogic and research purposes and an elaborated graphical user interface.

To remedy at these issues, we proposed an approach to develop a new simu- lation tool. We emphasized that this work would be realized in the framework of a collaborative and progressive project. In some sense, the work described in this document consists of the bootstrap of the project. It amounts the kernel of simulation development, in addition to some network protocols implementation.

Further in this document, we focused on the kernel of simulation achieve- ment steps. This second part detailed both analysis of requirements, design, and realization of the simulator kernel. First, in the analysis requirement chapter, we enumerated the expected functionalities and facilities from a network sim- ulation tool. Therefore, we release the different functional and non functional requirements. We illustrated with some use cases diagrams conform to the UML formalism.

Next, we moved to the kernel of simulation design description. We started with a general classes diagram overview. Then we described in details the most

Design and Implementation of a Wireless Networks Simulator 8.0

important classes. We concluded with two sequence diagrams that show in a simplified way the expected mode of functioning of the kernel, illustrating interactions between different classes.

After that, we gave a glance at the project realization hardware and software environment. Since the project remains in progress, we specified the stage of advancement we reached. As the project is collaborative, we mentioned the tasks and parts of the project that are currently in development by other collaborators.

All along the second part we insisted on the extreme modularity and ex- tensibility of our simulator. To prove this, we explained how to add a new protocol to the protocols library of the simulator with a concrete example. In deed, we implemented the IEEE standard protocol 802.11i. This was the aim of the third part of the report. We started with the description of the standard specifications on the protocol. Further, we explained how we designed it to be integrated in the simulator.

We tried to be as clear as possible in such way to allow anyone who reads the report using or extending the new simulation tool, as soon as it will be under OpenGL licence. This may explain the relative length of the document.

Since it is about a big project that surpasses the framework of this End of Studies Project, a lot of enhancements are still to be done. We distinguish immediate enhancements and long term enhancement.

Immediate enhancements are often closely related to the kernel of simula- tion. They sum up to some local modifications with no effects on other parts of the software. Among them, we cite the optimization of the calendar data structure for a better management of CPU resource. We cite also the addition of new classes that will take care of random variable, distribution and process generation. Another near future aim is that of developing an elaborated graphi- cal interface, taking the pattern of SimulX. Of course, making mobility features functional is also one of our primary priorities.

Concerning the long term perspectives, we can consider that the continuity and the success of the project depend on the upgrade of new protocols and the enrichment of the simulator protocol library. However, the addition of some protocols urges (e.g. IP, TCP, 802.11, 802.3). Indeed, these protocols represent the background of most simulation scenarios.

Personally, I particularly appreciated working on this project. It was a very enriching scientific experience as much as it was a good exercise of relatively big project design. This allowed me to set in practice some of the knowledge acquired during my course of engineering studies. All the more, it was a unique occasion to experiment participation in a collaborative project.

Therefore, I plan to continue working on the development of this simulator, but also to use it to simulate and evaluate new protocols, particularly in the framework of master thesis.

# Bibliography

1. B. Aboba and P. Calhoun. Radius (remote authentication dial in user ser- vice) support for extensible authentication protocol (eap). Technical Re- port RFC 3579, Internet Engineering Task Force. Network Working Group, 2003.
2. B. Aboba and D. Simon. Ppp eap tls authentication protocol. Techni- cal Report RFC 2716, Internet Engineering Task Force. Network Working Group, 1999.
3. L. Blunk and J. Vollbrecht. Ppp extensible authentication protocol (eap). Technical Report RFC 2284, Internet Engineering Task Force. Network Working Group, 1998.
4. A. Rubens C. Rigney, S. Willens and W. Simpson. Remote authentica- tion dial in user service (radius). Technical Report RFC 2865, Internet Engineering Task Force. Network Working Group, 2000.
5. Changhua He and John C Mitchell. Security analysis and improvements for ieee 802.11i. 2004.
6. IEEE. *IEEE Std 802.1X, 2001; IEEE Standard for Local and metropolitan area networks, Port Based Network Access Control*. 2001.
7. IEEE. *IEEE Std 802.11i 2004; IEEE Standard for Information technology, Telecommunications and information exchange between systems, Local and metropolitan area networks, Specific requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, Amendment 6: Medium Access Control (MAC) Security Enhancements*. 2004.
8. Ricki G. Ingalls. Introduction to simulation. *Simulation Conference, 2002. Proceedings of the Winter*, vol.1:7–16, 2002.
9. Raj Jain. *The Art of Computer Systems Performance Analysis*. John Wiley & Sons, 2001.
10. Romain Kuntz. Implementation du protocole mobile ipv6 dans un sim- ulateur de reseaux 802.11b. Technical report, Universite Louis Pasteur Strasbourg & LSiiT, 2003.

Design and Implementation of a Wireless Networks Simulator 8.0

1. Jean-Marie Bonnin Mohamed Kassab, Abdelfettah Belghith. Implementa- tion de methodes d authentification rapides sur un r´eseau ieee 802.11.
2. Jean-Marie Bonnin Sahbi Sassi Mohamed Kassab, Abdelfettah Belghith. Fast pre-authentication based on proactive key distribution for 802.11 in- frastructure networks.
3. N. Montavont and T. Noel. Anticipated handover over ieee 802.11 net- works. In *Wireless And Mobile Computing, Networking And Communica- tions, 2005. (WiMob’2005), IEEE International Conference on*, volume 2, pages 64–71Vol.2, 22-24 Aug. 2005.
4. A. Smith G. Zorn P. Congdon, B. Aboba and J. Roese. Ieee 802.1x remote authentication dial in user service (radius) usage guidelines. Technical Re- port RFC 3580, Internet Engineering Task Force. Network Working Group, 2003.
5. Robert E. Shannon. *Systems Simulation :The Art and Science*. Prentice Hall, 1975.
6. A. Varga. Using the omnet++ discrete event simulation system in educa- tion. *Education, IEEE Transactions on*, 42(4):11pp., 1999.
7. Dr.-Ing. Andreas Willig. Performance evaluation techniques of simulation software. Technical report, Telecommunication Networks Group (TKN) Technical University Berlin, Summer 2005.
8. Gregory K. Egan Y. Ahmet Sekercioglu, Andras Vagra. Parallel simulation made easy with omnet++. *European Simulation Symposium and Exhibi- tion*, 2003.