**LITERATURE SURVEY**

# 1) Smart city with Chinese characteristics against the background of big data: Idea, action and risk

**Authors:** [Yuzhe Wu](https://www.sciencedirect.com/author/35772932700/yuzhe-wu)**,**  Weiwen Zhang**.**

Chinese urbanization has generated great impacts on the world since the reform and opening up. However, urban problems, e.g., [environmental pollution](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/environmental-pollution), resources shortage, and traffic jam, have been more and more serious for urban management and development. Smart city has been put forward as an effective approach to achieve better urban management recently. Smart city aims to realize the integration of municipal service, business, transportation, water, energy source and other urban sub-systems through close combination of human wisdom and information communication techniques (ICTs). As a result, the link and synergy of information could be ultimately established with ICTs, e.g., internet, [internet of things](https://www.sciencedirect.com/topics/engineering/internet-of-things), cloud computing. Yet, few studies have been conducted to systematically link smart city with big data in China. This paper aims to put forward a development framework of smart city with Chinese characteristics against the background of big data. Key actions, including rational planning of city infrastructures, the establishment and improvement of long-acting mechanism, the effective performance of city managerial function, are proposed to realize the development idea. Meanwhile, this paper also investigates the risks embedded in development of smart city with Chinese characteristics, e.g., information safety, weak emergency responding capacity and poor independent research and development capacity of core [technology](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/science-and-technology). This study can facilitate Chinese local governments to systematically plan smart city before clinging the hot concept in a rush.

# 2.) Survey of improving Knearest-neighbor for classification

**AUTHORS:** [Zhihua Cai](https://ieeexplore.ieee.org/author/37272968800), [Dianhong Wang](https://ieeexplore.ieee.org/author/37310166300), and [Siwei Jiang](https://ieeexplore.ieee.org/author/37085592735)Gippsland School of Information Technology, Monash University, Churchil

KNN (k-nearest-neighbor) has been widely used as an effective classification model. In this paper, we summarize three main shortcomings confronting KNN and single out three main methods for overcoming its three shortcomings. Keeping to these methods, we try our best to survey some improved algorithms and experimentally tested their effectiveness. Besides, we discuss some directions for future study on KNN.

# 3. ) Using Machine Learning Algorithms to Analyze Crime Data

# 

**AUTHORS :** [**Lawrence McClendon**](https://www.researchgate.net/scientific-contributions/Lawrence-McClendon-2071943946?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIiwicGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIn19)[,](https://arxiv.org/search/cs?searchtype=author&query=Roth%2C+P)[**Natarajan Meghanathan**](https://www.researchgate.net/profile/Natarajan-Meghanathan?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIiwicGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIn19)

Data mining and machine learning have become a vital part of crime detection and prevention. In this research, we use WEKA, an open source data mining software, to conduct a comparative study between the violent crime patterns from the Communities and Crime Unnormalized Dataset provided by the University of California-Irvine repository and actual crime statistical data for the state of Mississippi that has been provided by neighborhoodscout.com. We implemented the Linear Regression, Additive Regression, and Decision Stump algorithms using the same finite set of features, on the Communities and Crime Dataset. Overall, the linear regression algorithm performed the best among the three selected algorithms. The scope of this project is to prove how effective and accurate the machine learning algorithms used in data mining analysis can be at predicting violent crime patterns.

# 4. ) Crime Prediction Using Decision Tree (J48) Classification Algorithm

**AUTHORS:** [**Emmanuel Ahishakiye**](https://www.researchgate.net/profile/Emmanuel-Ahishakiye-2?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIiwicGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIn19)**,** [**Danison Taremwa**](https://www.researchgate.net/scientific-contributions/Danison-Taremwa-2126606141?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIiwicGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIn19)

**,** [**Elisha Opiyo**](https://www.researchgate.net/profile/Elisha-Opiyo?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIiwicGFnZSI6InB1YmxpY2F0aW9uRGV0YWlsIn19)

With the growing processing power of computing systems and the increasing availability of massive datasets, machine learning algorithms have led to major breakthroughs in many different areas. This development has influenced computer security, spawning a series of work on learning-based security systems, such as for malware detection, vulnerability discovery, and binary code analysis. Despite great potential, machine learning in security is prone to subtle pitfalls that undermine its performance and render learning-based systems potentially unsuitable for security tasks and practical deployment. In this paper, we look at this problem with critical eyes. First, we identify common pitfalls in the design, implementation, and evaluation of learning-based security systems. We conduct a study of 30 papers from top-tier security conferences within the past 10 years, confirming that these pitfalls are widespread in the current security literature. In an empirical analysis, we further demonstrate how individual pitfalls can lead to unrealistic performance and interpretations, obstructing the understanding of the security problem at hand. As a remedy, we propose actionable recommendations to support researchers in avoiding or mitigating the pitfalls where possible. Furthermore, we identify open problems when applying machine learning in security and provide directions for further research.