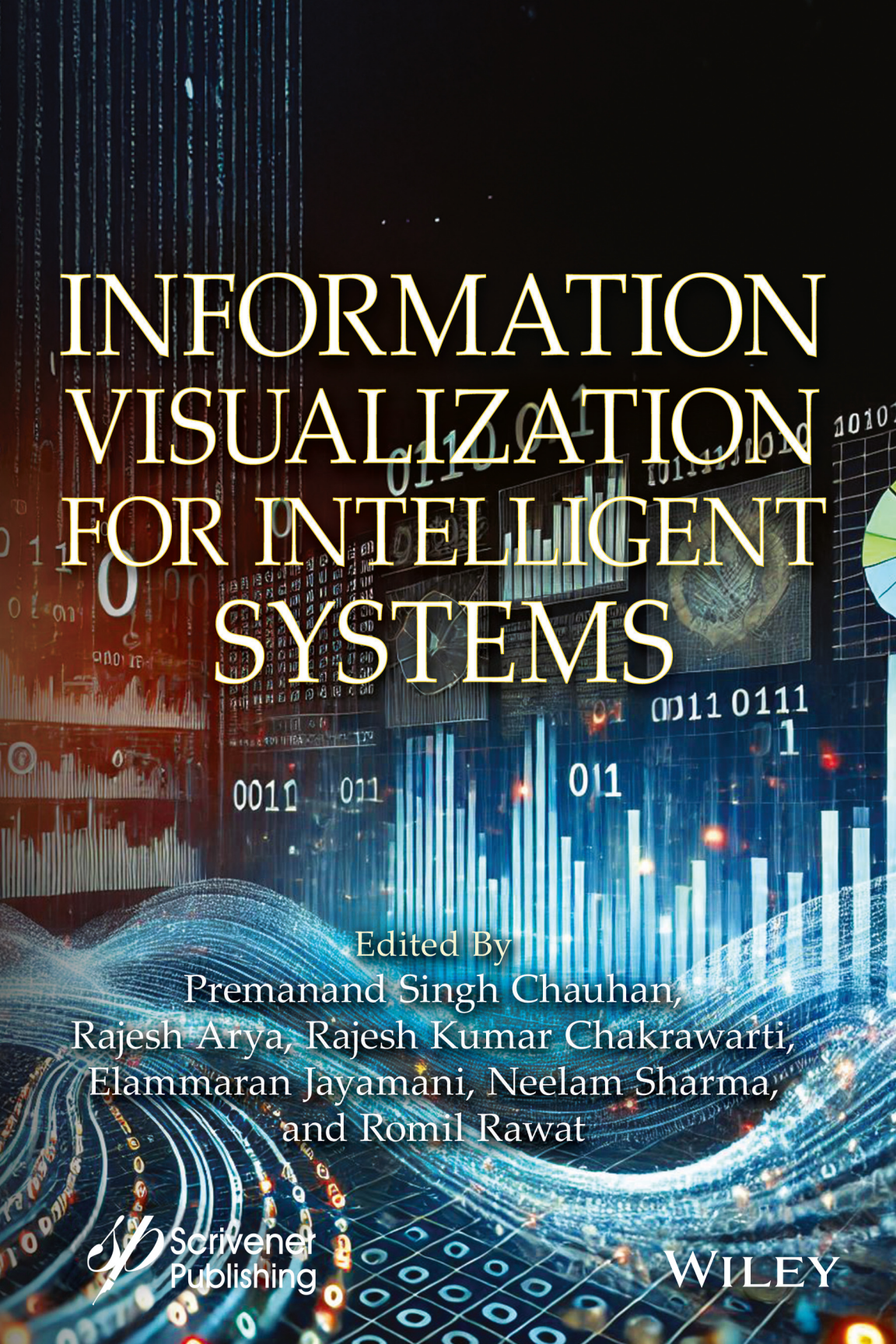
INFORMATION VISUALIZATION FOR INTELLIGENT SYSTEMS
Edited By
Premanand Singh Chauhan, Rajesh Arya, Rajesh Kumar Chakrawarti, Elammaran Jayamani, Neelam Sharma, and Romil Rawat
Scrivener Publishing
WILEY

Information Visualization for Intelligent Systems

Scrivener Publishing
100 Cummings Center, Suite 541J
Beverly, MA 01915-6106

Publishers at Scrivener
Martin Scrivener (martin@scrivenerpublishing.com)
Phillip Carmical (pcarmical@scrivenerpublishing.com)

Information Visualization for Intelligent Systems

Edited by

Premanand Singh Chauhan

Rajesh Arya

Rajesh Kumar Chakrawarti

Elammaran Jayamani

Neelam Sharma

and

Romil Rawat

WILEY

This edition first published 2025 by John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA and Scrivener Publishing LLC, 100 Cummings Center, Suite 541J, Beverly, MA 01915, USA

For more information about Scrivener publications please visit www.scrivenerpublishing.com.

Wiley Global Headquarters
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

ISBN 978-1-394-30578-0

Front cover images supplied by Adobe Firefly
Cover design by Russell Richardson

Set in size of 11pt and Minion Pro by Manila Typesetting Company, Makati, Philippines

Printed in the USA

10 9 8 7 6 5 4 3 2 1

Contents

Preface

The book focusses on advanced computing, or machine intelligence, the ability of a technology (a machine, device, or algorithm) to interact with its surroundings intelligently. This means that the technology can make decisions and take actions that will increase the likelihood that its objectives will be met. In contrast to the natural intelligence exhibited by people, artificial intelligence (AI), sometimes referred to as machine intelligence, is intelligence manifested by machines. The modern world is experiencing a period of paradigm shifts. New technologies have contributed to these shifts in part because they offer high-speed computing capabilities that make complicated machine intelligence systems possible. These advancements are paving the way for the creation of new cyber systems, which employ continually generated data to construct machine intelligence models that carry out specific functions inside the system. While the isolated use of cyber systems is becoming more common, the synchronic integration of these systems with other cyber systems to create a compact and intelligent structure that can interact deeply and independently with a physical system is still largely unanswered and has only been briefly discussed from a philosophical perspective in a few works.

Modern civilisation has undergone many paradigms shifts as a result of technological breakthroughs. These developments brought in immense data creation, cloud data storage systems, near-instantaneous worldwide information exchange, very quick computer capabilities, etc. Additionally, they paved the way for the development of cutting-edge cyber systems that employ systematically created data pipelines to carry out certain tasks. For instance, in certain nations, video surveillance imagery is used to detect criminals or possible criminals using machine intelligence (MI) models. Moreover, autonomous MI systems have applications in the medical field, where they enable prompt detection of infections like COVID-19. The chemical industry also uses a variety of applications.

1

Analysis of Restaurant Reviews Using Novel Hybrid Approach Algorithm Over Convolutional Neural Network Algorithm with Improved Accuracy

K. Abhilash Reddy and Uma Priyadarsini P.S.*

Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India

Abstract

The goal of this endeavor is to assess restaurant evaluations using a novel hybrid approach method in conjunction with the algorithm known as convolutional neural network (CNN). The study presents a novel hybrid approach that uses deep learning to classify restaurant evaluations as either good or negative. A collection of reviews was compiled in order to assess the efficacy of the proposed method. The hybrid approach algorithm (accuracy = 96.1%) analyzes the reviews and increases the measured accuracy over CNN (accuracy = 92.30%) with a statistically significant value of 0.004. The findings from the assessment conducted on the test dataset indicate that, in comparison to alternative methodologies currently in use, the novel hybrid approach technique yields the most precise reviews.

***Keywords*:** Sentiment analysis, novel hybrid approach, convolutional neural network, deep learning, restaurant reviews, polarity, food

Introduction

In recent years, the growth of restaurants through online platforms has been significant, with websites becoming the primary way for customers

**Corresponding author*: umapriyadarsini@saveetha.com

Premanand Singh Chauhan, Rajesh Arya, Rajesh Kumar Chakrawarti, Elammaran Jayamani, Neelam Sharma and Romil Rawat (eds.) Information Visualization for Intelligent Systems, (1–16)

to provide their opinions and assess the quality of restaurants and food services. The sentiment of customers can be inferred from these online reviews, which play a crucial role in shaping the reputation of an eatery. The potential to evaluate customers' opinions and make adjustments is provided by the contact between customers and owners via Internet platforms. Training machines with labeled data provides the benefit of more accurate future analysis of customer sentiment (Young *et al.*, 2018). The significance of this study extracted features from reviews and predicted their sentiment using a mixed deep learning technique. The findings of this study will help business owners by offering insightful information for making decisions. The procedure entails removing text from the web, classifying it, and figuring out how it feels. This contribution comprises a dataset comprising one thousand reviews sourced from Bangladesh. Chinsha T.C. proposed a feature based on an analysis of restaurant reviews (Chinsha and Joseph, 2015). The applications of the analysis of restaurant reviews using the novel hybrid approach algorithm include (Sharif, Hoque, and Hossain, 2019) the following: Improving Restaurant Operations: The analysis of customer feedback can be used by restaurants to identify areas for improvement and make informed decisions about menu offerings, food services, and other aspects of their operations. Customer Segmentation: Customers can be divided into groups according to their opinions and preferences using the hybrid approach, allowing restaurants to tailor their customers (Mohammad, Kiritchenko, and Zhu, 2013).

The examination of restaurant review systems has generated a significant amount of scholarly interest in recent years. On Google Scholar, over 191 papers were published, whereas on IEEE, over 97 papers were published. This research aims to analyze individuals' perspectives about restaurants using an innovative method. By utilizing deep learning architectures, this study seeks to achieve higher accuracy in sentiment analysis of restaurant reviews. The innovative combination of a convolutional neural network (CNN) and a novel hybrid strategy is suggested to address the diversity of recent datasets. A sample of 1000 reviews was collected and preprocessed to structure the unstructured and unlabeled data, with labels assigned as positive or negative. The CNN model learns the representation of words, while the novel hybrid approach learns more nuanced representations specific to the classification task. Hyperparameters were optimized before training the model to improve accuracy. Sentiment analysis is a common approach to predict customer reviews, as shown in previous studies by authors like Gan, who assessed restaurants based on factors like the quality of the food, cost, service, or atmosphere, and its context, and Jia, who created a model for categorizing restaurant reviews (Jia, 2020). Using a set of data of 1060

reviews, author Niphat Claypo developed a sentiment analysis framework using a combination of K-means clustering and the MRF feature selection (Claypo and Jaiyen, 2015). The optimal average K-means accuracy was 71.73%. An opinion mining model with 70% accuracy was proposed by author Sun using a dependency parser and Sentiwordnet (Sun, Luo, and Chen, 2017). Author Soujanya Poria developed the CNN methodology for identifying sentiment meaning using aspect uprooting. A multilayer CNN was used for word tagging according to Young *et al.* (2018).

The research gap is to overcome these limitations, a hybrid approach combining deep learning and sentiment analysis is proposed to classify sentiment polarity as either positive or negative. However, it is important to note that the accuracy of the machine learning models is contingent upon the quality of the training data, and any biases or inaccuracies could result in incorrect predictions. Additionally, the interpretability of some machine learning models, such as deep learning networks, can be limited, making it challenging to comprehend their reasoning. This study evaluates the effectiveness of a new hybrid approach method in comparison to a CNN algorithm for improving the accuracy of emotion polarity categorization in restaurant reviews.

Related Work

In the past five years, between 2017 and 2021, there has been an analysis of restaurant review systems that have been the subject of a sizable number of research articles. There are over 191 papers published on Google Scholar and over 97 papers published in IEEE. Hossain, Sharif, and Hoque (2020) conducted an analysis utilizing 4000 Bengali movie reviews, suggested a sentiment analysis model, and achieved a precision of 88.90% for long short-term memory (LSTM) and 82.42% for SVM. Utilizing the multinomial naive Bayes system that has an accuracy rating of 84%, 2000 Bengali critiques of books were utilized to categorize sentiment opposites into positive and negative categories (Hossein, Hoque, and Sarker, 2021). Sarker (2019) offered an LSTM-based sentiment analysis with an accuracy of 55.23% to categorize 1500 tweets into positive, negative, and neutral groups. The study of Wahid, Hasan, and Alom (2019) introduces a method for sentiment analysis utilizing LSTM (long short-term memory) on a dataset of 10,000 comments on Facebook to divide Bengali content into either positive or negative groups with 95% accuracy.

An LSTM-based algorithm was used to classify attitudes from Facebook tweets, achieving an 85% accuracy on a set of 10,000 Bengali messages.

Previous studies on sentiment analysis in Bengali focused on datasets such as Twitter posts, book reviews, and movie ratings, but these datasets were generally small. There is a dearth of research on sentiment analysis of restaurant reviews in the Bengali language other than this study. A total of 6625 restaurant reviews were gathered for the current study from a variety of online sources, including restaurant pages (1763), groups (1940), and Facebook comments (2922). Furthermore, 2000 restaurant reviews were acquired using the Yelp dataset. Data obtained from February 2020 to June 2020 contained inconsistent reviews. To address this, a filter was designed to exclude duplicates, comments with a minimum of three terms, mixed language evaluations, neutral sentiment evaluations, and reviews containing punctuation, numerals, and emojis. The filter produced a refined dataset of 6435 evaluations, which had been manually annotated by three annotators with 12 to 18 months of expertise in natural language processing (NLP). The annotation process entailed preserving the labels of 2000 evaluations from the Yelp dataset. Mohammad, Kiritchenko, and Zhu (2013) employed Cohen's Kappa to assess the inter-rater agreement among annotators for evaluating the annotation quality. The data exhibit good quality, as indicated by the average Kappa value of 0.81 (Kwok & Yu, 2013).

Existing Methodology

Convolutional Neural Network Algorithm

CNNs are a deep learning technique known for their effectiveness in image identification and may also be applied to text classification tasks like sentiment evaluation of restaurant reviews. In the case of restaurant reviews, the text data are transformed into numerical data using techniques such as tokenization, padding, and one-hot encoding. The CNN model is then trained using the numerical data. During training, the model learns the patterns and relationships between the words and phrases in the reviews and the corresponding sentiment (positive, neutral, or negative). After the model has undergone training, it can be utilized to categorize fresh reviews and forecast the sentiment expressed in the review. Evaluating the model's accuracy can be done by utilizing metrics such as precision, recall, and F1-score.

Overall, using a CNN algorithm for sentiment analysis of restaurant reviews can lead to effective and efficient classification results.

Algorithm Steps

Data preprocessing
1. Load the restaurant review dataset
2. Clean and preprocess the text data
3. Tokenize the text data into sequences of words or phrases
4. Pad the sequences to a fixed length
5. One-hot encode the sequences

Model building and training
6. Define the CNN architecture (number of layers, filter size, etc.)
7. Compile the model by defining the loss function, optimizer, and metrics
8. Train the model on the preprocessed data using a suitable batch size and number of epochs

Model evaluation
9. Evaluate the trained model on a test set of reviews
10. Calculate evaluation metrics such as precision, recall, and F1-score

Model deployment
11. Save the trained model for later deployment
12. Load the saved model and use it to classify new restaurant reviews and predict their sentiment.

Proposed Methodology

The dataset used in this research work is based on the concept of restaurant reviews (Govindarajan, 2014). The dataset used in this research is collected from Kaggle. The sample size was determined through GPower software, where two groups of 10 sets were selected. Using the GPower 3.1 tool, a pre-test strength value was determined with $\alpha = 0.05$ and power = 0.80, the necessary parameters for a test of statistical significance comparing two independent means. The study employed a novel hybrid method and CNN algorithm by utilizing Technical Analysis software. Human or animal samples were not used due to the absence of ethical approval requirements. The hardware configuration comprised an Intel i5 core processor paired with 16 GB of RAM. The software utilized included HTML, Python, Java, Tomcat/Glassfish server, Jupyter Notebook, My SQL database, CSS web technologies, and J2SDK1.5 Java version.

Novel Hybrid Approach Algorithm

The novel hybrid approach algorithm for analyzing restaurant reviews combines multiple techniques from NLP to provide a comprehensive analysis of customer opinions and sentiments. The algorithm uses methods such as sentiment analysis, topic modeling, and entity recognition to identify trends and patterns in the data. The objective is to help restaurant owners and managers make informed decisions about their business and food services by understanding the strengths and weaknesses of their establishment through the analysis of customer reviews. The insights obtained from the study can be used to improve the customer experience and enhance overall satisfaction.

Pseudocode for a Hybrid Approach Algorithm for Analyzing Restaurant Reviews:

Step 1: Preprocessing
- Input: Raw restaurant reviews
- Output: Cleaned and preprocessed reviews
- Remove irrelevant information such as punctuation, stop words, and special characters.
- Tokenize the reviews into individual words.
- Utilize lemmatization or stemming to simplify words to their most basic form.

Step 2: Sentiment Analysis
- Input: Cleaned and preprocessed reviews
- Output: Sentiment scores for each review
- Utilize an analysis of sentiment tool or algorithm to categorize each review as favorable, negative, or neutral.
- Calculate the sentiment score for each review based on the classification results.

Step 3: Topic Modeling
- Input: Cleaned and preprocessed reviews
- Output: Topics and their distributions in the reviews
- Apply LDA (latent Dirichlet allocation) for topic modeling to determine the primary subjects covered in the reviews.
- Calculate the distribution of topics in the reviews.

Step 4: Entity Recognition
Input: Cleaned and preprocessed reviews
Output: Entities mentioned in the reviews
Use an entity recognition tool or model to identify entities such as food items, services, ambiance, etc. mentioned in the reviews.
Store the entities and their mentions in a data structure.

Step 5: Combination of Results
Input: Sentiment scores, topic distributions, and entity mentions
Output: comprehensive analysis of the restaurant reviews
Combine the results from the sentiment analysis, topic modeling, and entity recognition steps to obtain a comprehensive analysis of the restaurant reviews.
Use the results to identify patterns and trends in the data.

Step 6: Visualization and Reporting
Input: Comprehensive analysis of the restaurant reviews
Output: Visualizations and reports
Use data visualization techniques to present the results of the analysis in an intuitive and easy-to-understand manner.

Statistical Analysis

In order to statistically analyze restaurant evaluations, SPSS software employs a special hybrid approach strategy that outperforms the naive Bayes approach in terms of accuracy. Accuracy is the dependent variable, and efficacy is the independent variable. A sample size of ten has been used for the analysis. The novel hybrid approach algorithm's accuracy is determined using the statistical T-test analysis.

Results

The hybrid approach algorithm and the CNN algorithm were run independently in Jupyter Notebook using an experiment size of 10.

Table 1.1 displays the expected accuracy and loss of the hybrid approach algorithm.

Table 1.2 represents the CNN's expected accuracy and loss. The statistical values that can be utilized for comparison are computed for each of

Table 1.1 An analysis of accuracy and loss was conducted on a novel hybrid approach algorithm using a sample size of 10. The hybrid approach algorithm achieves a precision rate of 96.10% with a corresponding error rate of 3.90%.

Iteration	Accuracy (%)	Loss (%)
1	96.63	3.37
2	97.46	2.54
3	98.25	1.75
4	94.52	5.48
5	95.66	4.34
6	95.68	4.32
7	95.25	4.75
8	96.25	3.75
9	95.20	4.80
10	96.10	3.90

the 10 data samples together with the associated loss values. According to the findings, the CNN had a mean accuracy of 92.30% and the hybrid approach algorithm had a mean accuracy of 96.1%.

Table 1.3 represents the mean accuracy scores for the CNN and the hybrid approach algorithm. When compared to the CNN, the hybrid approach algorithm's mean value is superior, with standard deviations of 2.08555 and 2.18436, respectively.

Table 1.4 shows the hybrid approach algorithm and CNN-independent sample T-test data with a statistically significant value of 0.004 ($p < 0.05$).

Table 1.5 evaluates the convolutional neural network, also known as CNN [20–25], algorithm's accuracy against that of the hybrid approach algorithm. The hybrid approach algorithm attains an efficacy of 96.10%, whereas the CNN (convolutional neural network) algorithm yields a success rate of 92.30%. In terms of accuracy, the hybrid approach algorithm performs better than the CNN model algorithm.

Table 1.2 A CNN with a sample size of 10 is used to analyze its accuracy and loss. 92.30% precision and 7.70% error rates are attained using the CNN.

Iteration	Accuracy (%)	Loss (%)
1	91.15	9.85
2	94.32	5.68
3	93.46	6.54
4	91.07	9.93
5	93.18	6.82
6	94.74	5.26
7	90.56	9.44
8	90.78	9.22
9	91.44	8.56
10	92.30	7.70

Table 1.3 The group statistics for the novel hybrid model show a mean accuracy of 96.10% with a standard deviation of 1.11. In comparison, the LSTM model has a mean accuracy of 92.30% with a standard deviation of 1.52.

	Group	N	Mean	Std. deviation	Std. error of the mean
Accuracy	Hybrid algorithm	10	96.1000	1.11427	0.35236
	CNN	10	92.3000	1.52996	0.48382

Figure 1.1 illustrates the flow chart of a distinctive hybrid methodology that integrates a CNN (convolutional neural network) to analyze the sentiments of online restaurant reviews.

Figure 1.2 shows the average accuracy difference between the hybrid approach algorithm and the CNN. With a mean of 93.3571, a standard

Table 1.4 Performing a T-test on an independent sample to establish significance and compute the standard error. A p-value of less than 0.05 is considered statistically significant, and a confidence interval of 95% was calculated. The hybrid approach algorithm demonstrates superior performance compared to CNNs, with a statistically significant p-value of 0.004, meeting the criteria for significance at a two-tailed level of $p < 0.05$.

Levene's equality of variances test				**T-test for means equality**					**95% Confidence interval for the difference**	
		F	**Sig**	**t**	**df**	**Sig. (two-tailed)**	**Mean difference**	**Std. error difference**	**Lower**	**Upper**
Accuracy	**Assuming equal variance**	0.309	0.014	6.349	18	0.004	3.80000	0.59853	2.54254	5.05746
	No assumption of equal variances			6.349	16.451	0.004	3.80000	0.59853	2.53400	5.06600