Edusaintek: Jurnal Pendidikan, Sains dan Teknologi
Volume 9 Issue 3 2022 Pages 920-929
p-ISSN: 1858-005X e-ISSN: 2655-3392 DOI:https://doi.org/10.47668/edusaintek.v9i3.611
website: https://journalstkippgrisitubondo.ac.id/index.php/EDUSAINTEK

ANALISIS PERBANDINGAN AKURASI K-NEAREST NEIGHBOR DAN NAÏVE BAYES UNTUK KLASIFIKASI DATA SERANGAN JARINGAN KOMPUTER

Muhammad Iqbal[1*], Rd. Rohmat Saedudin[2], & Muhammad Fathinuddin[3]
[1,2,3]Universitas Telkom, Indonesia
*e-mail: emuhammadiqbal@student.telkomuniversity.ac.id

Abstract: There are many organizations and individuals who do not understand network security so that they get potential attacks and experience system damage. To prevent potential attacks, an Intrusion Detection System (IDS) was developed. Some of the non-machine learning methods used are not yet accurate, so they require a more accurate machine learning method to detect attacks. To solve the problem, in this study, a comparison is made using the K-Nearest Neighbor and Naïve Bayes methods to detect computer network attacks optimally. In this study, the implementation uses the K-Nearest Neighbor and Naïve Bayes methods in detecting HTTPDoS attacks using the ISCX testbed dataset on June 14, 2012 which consists of 157,867 packets and as many as 19 features. This study analyzes the comparison of methods that will result from the classification process with the confusion matrix and ROC curve. The final result of the research is that the KNN method produces an accuracy percentage of 99.994% and has a very good data classification quality compared to the Naïve Bayes accuracy percentage of 39.885%.

Keywords: *KNN, IDS, Nave Bayes, Classification, Attack*

Abstrak: Banyaknya organisasi maupun individu yang belum paham terhadap keamanan jaringan sehingga mendapatkan potensi serangan dan mengalami kerusakan sistem. Untuk melakukan pecegahan potensi serangan dikembangkan yaitu *Intrusion Detection System* (IDS). Dari beberapa metode non-machine learning yang digunakan belum akurat, sehingga memerlukan metode dengan machine learning yang lebih akurat untuk mendeteksi serangan. Untuk mengatasi permasalahan, dalam penelitian melakukan perbandingan menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes* untuk mendeteksi serangan jaringan komputer dengan optimal. Dalam penelitian ini, implementasi menggunakan metode K-Nearest Neighbor dan Naïve Bayes dalam mendeteksi serangan HTTPDoS dengan menggunakan dataset ISCX testbed 14 Juni 2012 yang terdiri dari 157.867 paket dan sebanyak 19 fitur. Penelitian ini menganalisis perbandingan metode yang akan dihasilkan dari proses klasifikasi dengan confusion matrix dan kurva ROC. Pada hasil akhir penelitian yang diperoleh adalah metode KNN menghasilkan persentase akurasi sebesar 99,994% dan memiliki kualitas klasifikasi data yang sangat baik dibandingkan persentase akurasi Naïve Bayes 39,885%.

Kata Kunci: *KNN, IDS, Naïve Bayes, Klasifikasi, Serangan*

PENDAHULUAN

Keamanan jaringan adalah hal yang penting untuk diperhatikan, terutama pada era teknologi saat ini (Nugroho et al., 2019; Zabar and Novianto, 2015).

Banyaknya organisasi maupun individu yang tidak peduli terhadap masalah keamanan yang dimiliki (Purba and Efendi, 2021). Sehingga ketika jaringan mendapatkan serangan dan mengalami kerusakan sistem, disaat itu juga harus mengeluarkan biaya untuk melakukan perbaikan sistem yang dirusak (Triyansyah and Fitrianah, 2018). Informasi yang diperoleh dari Pusat Operasi Keamanan Operasi Keamanan Siber Nasional (Pusopskamsinas) Badan Siber dan Sandi Negara (BSSN) mencatat bahwa ada sekitar 88.414.296 serangan telah terjadi sejak 1 Januari hingga 12 April 2020 (Al Fikri and Djuniadi 2021; Marcus, Rosyadi, and Pamuji 2021; Panggabean 2018).

Jumlah serangan maksimum terjadi pada 12 Maret 2020 mencapai 3.344.470 serangan, kemudian jumlah serangan menurun secara signifikan ketika kebijakan *Work From Home* (WFH) diterapkan di berbagai tempat (Saputra, 2019). Untuk melakukan pecegahan terhadap potensi serangan sudah dikembangkan oleh suatu sistem atau metode yang dikenal dengan *Intrusion Detection System* (IDS). *Intrusion Detection System* (IDS) merupakan sebuah aplikasi perangkat lunak atau perangkat keras yang dapat mendeteksi aktivitas yang diduga mencurigakan pada sistem atau jaringan (Al Fikri and Djuniadi, 2021).

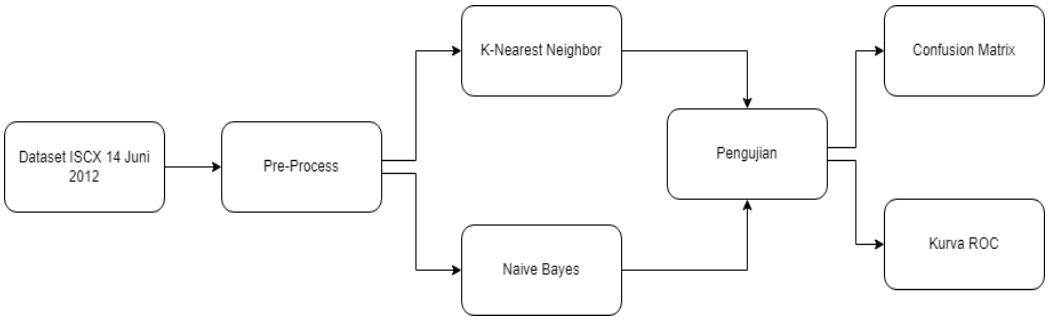
Dari beberapa penelitian yang telah dilakukan untuk membandingkan metode klasifikasi data serangan jaringan komputer. Penelitian yang dilakukan oleh Fibrianda and Bhawiyuga (2018) menyatakan bahwa kinerja algoritma Naive Bayes lebih baik dibanding dengan kinerja algoritma SVM. Penelitian kedua melakukan perbandingan dengan metode *K-Nearest Neighbor* dan *Decision Tree* oleh Ilham Ramadhan, Parman Sukarno dan Muhammad Arief Nugroho. Hasil dari penelitiannya menyatakan bahwa kinerja algoritma K-Nearest Neighbor lebih baik dibandingkan dengan kinerja algoritma *Decision Tree*. Terdapat penelitian terkait berikutnya yaitu tentang perbandingan nilai akurasi dari metode *Probabilistic Neural Network* dan *Naive Bayes*.

Kinerja dari algoritma Naïve Bayes lebih baik dibandingkan dengan algoritma *Probabilistic Neural Network* (Kusy and Kowalski 2018; Zeinali and Story 2017). Perbedaan penelitian ini dengan sebelumnya yaitu pada penelitian ini menganaliis perbandingan akurasi K-Nearest Neighbor dan Naïve Bayes untuk

klasifikasi data serangan jaringan komputer.

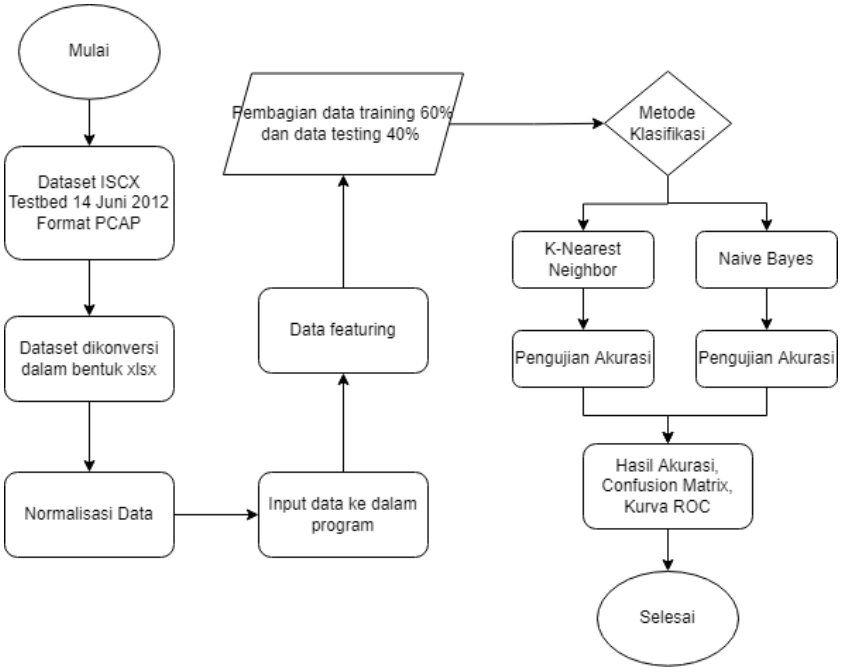
METODE

Metode penelitian yang digunakan adalah kualitatif dengan desain membandingkan antara dua alogaritma yang berbeda kemudian menemukan algoritma yang paling baik untuk tingkat akurasinya dalam mengklasifikasikan dari suatu dataset (Sugiyono, 2018; Sugiono, 2019). Penulis mengidentifikasi tiga langkah umum untuk pemecahan masalah yang sistematis. Pada tahap awal penulis melakukan identifikasi masalah berdasarkan pada studi kasus yang akan diteliti. Penulis mendefinisikan beberapa masalah diantaranya bagaimana akurasi dari K-Nearest Neighbor dan Naïve Bayes untuk klasifikasi data serangan jaringan komputer dan bagaimana perbandingan akurasi dari kedua algoritma yang digunakan. Selanjutnya menentukan tujuan diantaranya hasil akurasi dari K-Nearest Neighbor dan Naïve Bayes dan perbandingan hasil dari kedua algoritma yang digunakan yang dijelaskan pada gambar 1.



Gambar 1. Perancangan Lingkungan Pengujian

Kemudian pada proses terakhir dilanjutkan dengan melakukan pemilihan dataset yang akan digunakan didalam penelitian ini. Pada tahap pengumpulan data ini menggunakan dataset ISCX dan menggunakan dua metode evaluasi, diantaranya Uji Dependabilitas dan Uji Konfirmabilitas/Objektivitas. Terkait teknik pengolahan dijelaskan pada gambar 2.



Gambar 2. Pengolahan Data

HASIL DAN PEMBAHASAN

Hasil Confusion Matrix

Pada perhitungan confusion matrix menampilkan jenis paket, jumlah paket yang terdeteksi, accuracy, precision, recall dan f1 score. Lalu jenis paket terdiri diantaranya True-Negative, False-Positive, False-Negative, dan True-Positive. Jumlah paket adalah jumlah data yang telah dideteksi oleh setiap jenis paket (Caelen, 2017; Zeng, 2020).

Pada hasil perhitungan confusion matrix yang telah dilakukan pada saat proses klasifikasi dengan metode K-Nearest Neighbor dapat di jabarkan hasil yang didapat dijelaskan dibawah ini:

```
[23] acc_knn = accuracy_score(y_test, y_pred) #akurasi knn
     prec_knn = precision_score(y_test, y_pred) #presisi knn
     rec_knn = recall_score(y_test, y_pred) #recall knn
     f1_score_knn = 2*(prec_knn*rec_knn)/(prec_knn+rec_knn) #f1 score knn

     #menampilkan hasil evaluasi model
     print('Skor Akurasi kNN:' , acc_knn)
     print('Skor Presisi kNN:' , prec_knn)
     print('Skor Recall kNN:' , rec_knn)
     print('Skor F1 Score kNN:', f1_score_knn)

     Skor Akurasi kNN: 0.99995
     Skor Presisi kNN: 1.0
     Skor Recall kNN: 0.9999499949995
     Skor F1 Score kNN: 0.9999749968746093
```

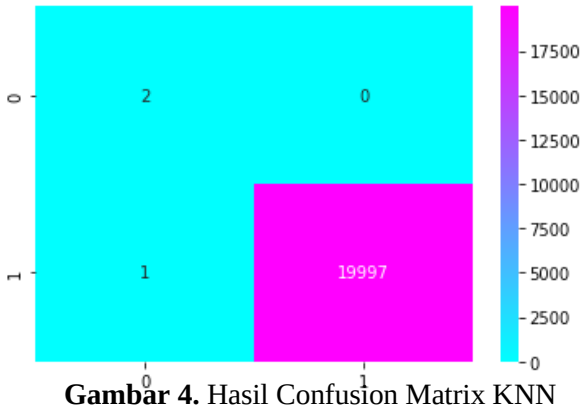
Gambar 3. Script Accuracy, Precision, Recall, F1 Score KN

Pada tabel 1, disajikan hasilkan dari masing-masing nilai accuracy, precision, recall, dan juga f1 score dari modul classify K-Nearest Neighbor seperti berikut ini:

Tabel 1. Perhitungan Confusion Matrix K-Nearest Neighbor

Jenis Paket	**Jumlah Paket**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
True Negative	2	99,995%	100%	99,994%	
False Positive	0				
False Negative	1				99,997%
True Positive	19997				

Pada tabel 1 disajikan hasil dari perhitungan confusion matrix menggunakan metode K-Nearest Neighbor dengan menunjukkan jumlah paket yang terdeteksi sebagai True-Negative sebanyak 2 data, False-Positive sebanyak 0 data, False-Negative sebanyak 2 data dan True-Positive sebanyak 19997 data, menghasilkan nilai accuracy sebesar 99,995%, precision sebesar 100%, recall sebesar 99,994% dan F1 score sebesar 99,997%. Dari nilai yang dihasilkan pada tabel diatas menunjukkan bahwa memiliki arti yaitu perbandingan dari setiap data yang ada, kemudian data yang ditentukan tersebut dapat dipastikan benar-benar merupakan data serangan atau data normal dari total accuracy data sebesar 99,995% benar. Selanjutnya dilihat dari hasil precision antara informasi yang diminta dengan jawaban yang diberikan oleh sistem memiliki persentase sebesar 100% sangat tepat. Lalu jika dilihat dari tingkat keberhasilan recall yaitu sebesar 99,994%. Dalam pengujian klasifikasi ini bisa dikatakan cukup berhasil untuk kualitasnya karena mencapai nilai precision dan recall yang cukup tinggi.



Gambar 4. Hasil Confusion Matrix KNN

Selanjutnya untuk hasil perhitungan confusion matrix yang dilakukan pada proses klasifikasi dengan metode Naïve Bayes, maka dapat dijabarkan hasilnya melalui penjelasan dibawah ini:

```
#mencari akurasi, presisi, recall, dan f1 score Naive Bayes
acc_nb = accuracy_score(y_test, y_pred_nb)
prec_nb = precision_score(y_test, y_pred_nb)
rec_nb = recall_score(y_test, y_pred_nb)
f1_score_nb = 2 * (prec_nb * rec_nb)/(prec_nb+rec_nb)
```

```
#menampilkan hasil evaluasi model
print('Skor Akurasi Naive Bayes:' , acc_nb)
print('Skor Presisi Naive Bayes:' , prec_nb)
print('Skor Recall Naive Bayes:' , rec_nb)
print('Skor Recall Naive Bayes:' , f1_score_nb)
```

```
Skor Akurasi Naive Bayes: 0.39885
Skor Presisi Naive Bayes: 1.0
Skor Recall Naive Bayes: 0.3987898789878988
Skor Recall Naive Bayes: 0.5701926858041684
```

Gambar 5. Script Accuracy, Precision, Recall, F1 Score Naive Bayes

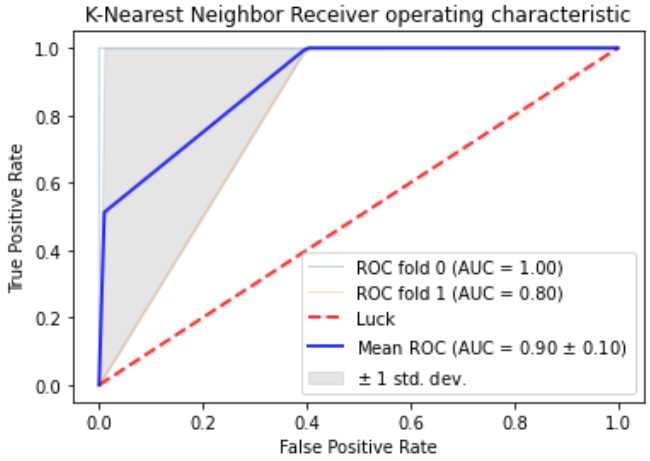
Pada tabel 2, disajikan hasilkan dari masing-masing nilai accuracy, precision, recall, dan juga f1 score dari modul classify Naïve Bayes seperti berikut ini:

Tabel 2. Perhitungan Confusion Matrix Naive Bayes

Jenis Paket	Jumlah Paket	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
True Negative	2	39,885%	100%	39,878%	
False Positive	0				
False Negative	12023				57,019%
True Positive	7975				

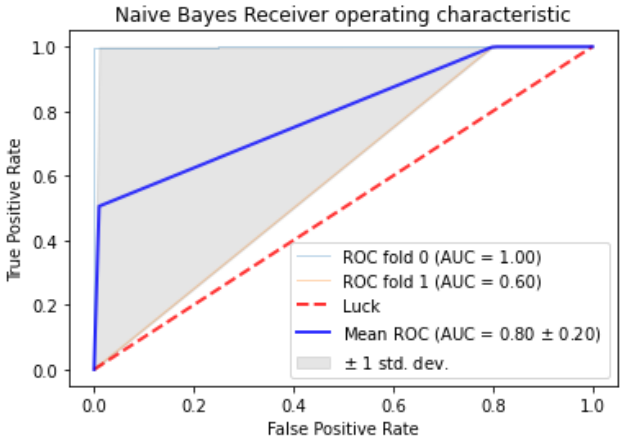
Pada tabel 2, disajikan hasil dari perhitungan confusion matrix dengan menggunakan metode Naïve Bayes dengan menunjukkan jumlah paket yang terdeteksi sebagai True-Negative sebanyak 2 data, False-Positive sebanyak 0 data, False-Negative sebanyak 12023 data dan True-Positive sebanyak 7975 data, menghasilkan nilai accuracy sebesar 39,885%, precision sebesar 100%, recall sebesar 39,878% dan F1 score sebesar 57,019%. Dari nilai yang dihasilkan pada tabel diatas menunjukkan bahwa memiliki arti yaitu perbandingan dari setiap data yang ada, kemudian data yang ditentukan tersebut dapat dipastikan benar-benar merupakan data serangan atau data normal dari total accuracy data sebesar 39,885% benar. Selanjutnya dilihat dari hasil precision antara informasi yang diminta dengan jawaban yang diberikan oleh sistem memiliki persentase sebesar 100% sangat tepat. Lalu jika dilihat dari tingkat keberhasilan recall yaitu sebesar 39,878%. Dalam pengujian klasifikasi ini untuk kualitasnya mencapai nilai precision tinggi dan recall yang rendah.

Berdasarkan hasil dari nilai True-Positive dan False-Positive yang didapatkan melalui perhitungan confusion matrix yang dilakukan selama klasifikasi pada modul classify dengan menggunakan metode K-Nearest Neighbor, menghasilkan kurva ROC pada gambar 6 seperti dibawah:



Gambar 6. Kurva ROC dengan Cross Validation KNN

Pada gambar 6 menjelaskan bahwa nilai dari AUC yang dihasilkan oleh kurva dari ROC K-Nearest Neighbor adalah sebesar 0,9 yang dimana memiliki arti bahwa nilai diagnostik termasuk kategori sangat baik dalam proses klasifikasi. Selanjutnya berdasar hasil dari nilai True-Positive dan False-Positive yang didapatkan melalui perhitungan confusion matrix yang dilakukan selama klasifikasi pada modul classify dengan menggunakan metode Naïve Bayes, menghasilkan kurva ROC pada gambar 7 seperti dibawah:



Gambar 7. Kurva ROC dengan Cross Validation Naive Bayes

Pada gambar 7 menjelaskan bahwa nilai dari AUC yang dihasilkan oleh kurva dari ROC Naïve Bayes adalah sebesar 0,8 yang dimana memiliki arti bahwa nilai diagnostik termasuk kategori sedang dalam proses klasifikasi.

SIMPULAN

Klasifikasi menggunakan metode K-Nearest Neighbor hasil dari confusion

matrix menghasilkan accuracy 99,995%, precision 100%, recall 99,994%, dan f1 score 99,997%. Dalam validasi teknik klasifikasi, penulis menggunakan kurva ROC dengan teknik Cross Validation, menghasilkan AUC sebesar 90% termasuk kedalam kategori sangat baik untuk klasifikasi. Klasifikasi menggunakan metode Naïve Bayes hasil dari confusion matrix menghasilkan accuracy 39,885%%, precision 100%, recall 39,885%, dan f1 score 57,019%.

Dalam validasi teknik klasifikasi, penulis menggunakan kurva ROC dengan teknik Cross Validation, menghasilkan AUC sebesar 80% berarti termasuk kedalam kategori baik untuk klasifikasi. Pada kedua classifier dikatakan sangat baik karena memiliki persentase precision yang tinggi, sedangkan recall pada classifier Naïve Bayes lebih rendah dibandingkan dengan classifier K-Nearest Neighbor yang lebih unggul. Berdasar hasil dari nilai AUC yang didapat K-Nearest Neighbor sebesar 90% sedangkan Naïve Bayes sebesar 80% dapat dikatakan bahwa K-Nearest Neighbor adalah algoritma yang lebih baik dalam melakukan klasifikasi data serangan jaringan komputer.

DAFTAR RUJUKAN

Caelen, Olivier. 2017. "A Bayesian Interpretation of the Confusion Matrix." *Annals of Mathematics and Artificial Intelligence* 81(3–4).

Fibrianda, Mercury Fluorida, and Adhitya Bhawiyuga. 2018. "Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM)." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* II(9).

Al Fikri, Khashaaisha, and Djuniadi. 2021. "Keamanan Jaringan Menggunakan Switch Port Security." *InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan* 5(2).

Kusy, Maciej, and Piotr A. Kowalski. 2018. "Weighted Probabilistic Neural Network." *Information Sciences* 430–431.

Marcus, Ronald David, Hudan Eka Rosyadi, and Fandi Yulian Pamuji. 2021. "Prototype Sistem Administrasi Dan Keamanan Jaringan Komputer Berbasis DHCP Server Mikrotik." *Briliant: Jurnal Riset dan Konseptual* 6(3).

Nugroho, Fendy Prasetyo, Robi Wariyanto Abdullah, Sri Wulandari, and Hanafi. 2019. "Keamanan Big Data Di Era Digital Di Indonesia." *Jurnal Informa* 5(1).

Panggabean, Parningotan. 2018. "Analisis Network Security Snort Metode Intrusion Detection System Untuk Optimasi Keamanan Jaringan Komputer." *Jursima* 6(1).

Purba, Winrou Wesley, and Rissal Efendi. 2021. "Perancangan Dan Analisis Sistem Keamanan Jaringan Komputer Menggunakan SNORT." *AITI* 17(2).

Saputra, D Dio Azmi. 2019. "Keamanan Jaringan Komputer." *Keamanan Jaringan Komputer*.

Sugiyono. 2018. *Metode Penelitian Kualitatif, Kuantitatif, Dan R&D*. CV.Alfabeta.

———. 2019. *Metode Penelitian Kuantitatif, Kualitatif, Dan R&D*. 1st ed. Bandung: Penerbit Alfabeta.

Triyansyah, D., and D. Fitrianah. 2018. "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing." *InComTech* 8(3): 163–82.

Zabar, Adzan Abdul, and Fahmi Novianto. 2015. "Keamanan Http Dan Https Berbasis Web Menggunakan Sistem Operasi Kali Linux." *Komputa : Jurnal Ilmiah Komputer dan Informatika* 4(2).

Zeinali, Yasha, and Brett A. Story. 2017. "Competitive Probabilistic Neural Network." *Integrated Computer-Aided Engineering* 24(2).

Zeng, Guoping. 2020. "On the Confusion Matrix in Credit Scoring and Its Analytical Properties." *Communications in Statistics - Theory and Methods* 49(9).