

Komparasi Metode Algoritma *Data Mining* pada Prediksi Uji Kelayakan *Credit Approval* pada Calon Nasabah Kredit Perbankan

Sari Dewi

Program Studi Sistem Informasi, Kampus Kota Pontianak
Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika
Jl. Abdurrahman Saleh No18 A Pontianak
E-mail :sari.sre@bsi.ac.id

ABSTRAKSI

Di era saat ini pemanfaatan data mining dalam menentukan keberhasilan meningkatkan layanan perbankan sangatlah efektif, salah satunya pemberian kredit kepada customer bank, menentukan credit approval memerlukan sistem yang akan digunakan untuk menjalankan proses pengajuan kredit serta didukung dengan kecermatan dalam memilih calon nasabah kredit yang baik sehingga dapat meminimalisir terjadinya kredit macet yang tidak diinginkan. Untuk mendukung hasil dari tingkat keberhasilan marketing dalam perannya untuk memasarkan produk layanan perbankan yang prosesnya membutuhkan data calon nasabah ini, maka dukungan data mining sangat berperan penting dalam klasifikasi calon nasabah bank yang akan mengambil kredit di bank. Berdasarkan pemetaan penelitian mengenai dukungan data mining pada calon nasabah didapat ada algoritma klasifikasi yang sering digunakan untuk klasifikasi calon nasabah antara lain Neural Network, Naive Bayes dan K-NN dalam prediksi keberhasilan marketing dalam menentukan kelayakan dari nasabah peminjam kredit bank dari uji coba yang dilakukan maka algoritma Neural Network lah yang lebih akurat dengan akurasi 90,71% dengan nilai AUC 0.880, hal ini dapat menjadi perbandingan data mining klasifikasi. Melihat nilai AUC dari ketiga metode tersebut yaitu NN, Naive Bayes dan K-NN, maka 3 algoritma tersebut termasuk kelompok klasifikasi baik karena nilai AUC-nya antara 0.80-1.00.

Kata kunci: *Data mining, Neural Network, Naive Bayes, K-NN, klasifikasi*

ABSTRACT

In the current era, the use of data mining in determining the success of improving banking services is very effective, one of which is lending to bank customers, determining credit approval requires a system that will be used to run the credit application process and supported by careful selection of prospective credit customers so that they can minimize the occurrence of bad credit. To support the results of the success rate of marketing in its role in marketing banking service products whose processes require prospective customer data, data mining support plays an important role in the classification of prospective bank customers who will take credit at the bank. classification algorithm that is often used for the classification of prospective customers, including NN, NB and K-NN in predicting marketing success in determining the feasibility of bank loan borrowers then the NN algorithm is more accurate with an accuracy of 90.71% with an AUC 0.880 value, this data mining classification comparison. Seeing the AUC value of the 3 methods namely NN, NB and K-NN, then the 3 algorithms are included in the group classification is good because the AUC value is between 0.80-1.00.

Keywords: *Maning data, Neural Network, Naive Bayes, K-NN, classification*

1. PENDAHULUAN

Pada dunia perbankan proses pemberian kredit kepada nasabah adalah kegiatan yang harus mempunyai standar uji kelayakan. Dalam pelaksanaannya, kredit yang bermasalah (kredit macet) sering terjadi akibat analisis kredit yang tidak efektif atau kurang cermat dalam proses pemberian kredit, maupun dari karakter nasabah yang tidak baik. Untuk mencegah terjadinya kredit macet, seorang analisis kredit perbankan harus mampu mengambil keputusan yang tepat untuk menerima ataupun menolak pengajuan kredit. Agar kredit yang diberikan mencapai sasaran, yaitu aman, maka analisis kredit dilakukan. Analisis kredit adalah kajian yang dilakukan untuk mengetahui kelayakan dari suatu permasalahan kredit. Melalui hasil analisis kreditnya, dapat diketahui apakah usaha nasabah layak (*feasible*) dan *marketable* (hasil usaha dapat dipasarkan), dan *profitable* (menguntungkan), serta dapat dilunasi tepat waktu. (Kusrini et al. 2009) Mengingat resiko tidak kembalinya kredit pada sebuah leasing, selalu ada, maka analisis kredit dengan cermat perlu dilakukan. Penilaian kuantitatif dan kualitatif dalam menganalisa kredit akan memberikan kejelasan bagi pembuat keputusan. Untuk mewujudkan hal tersebut, perlu dilakukan persiapan kredit, yaitu dengan mengumpulkan informasi dan data untuk bahan analisis. Kualitas hasil analisis tergantung pada kualitas SDM, data yang diperoleh, dan teknik analisis. Dengan menggunakan teknologi di bidang data mining yang mengoptimasi proses pencarian informasi prediksi dalam basis data yang besar, serta menemukan pola-pola yang tidak diketahui sebelumnya. Maka identifikasi pola data dari sistem penentuan pemberian kredit dapat dilakukan dengan pendekatan probabilitas bersyarat. Dukungan data

mining pada pemasaran adalah pada Approval Kredit. Dalam mengoptimisasi proses uji kelayakan nasabah diperlukan suatu strategi sehingga dapat digunakan untuk mencegah kredit macet, *Data mining* dalam uji kelayakan nasabah kredit menggunakan salah satunya menggunakan data calon nasabah

untuk melakukan proses pencarian pengetahuan baru guna mendukung pengambilan keputusan, Oleh karena itu, penelitian ini fokus pada pemanfaatan *data mining* untuk membandingkan tingkat kelayakan aproval kredit calon nasabah bank dari berbagai produk layanan perbankan, sehingga dapat diketahui apakah calon nasabah yang bersangkutan merupakan nasabah yang berpotensi menjadi nasabah kredit yang produktif atau tidak di liat dari penelitian sebelumnya algoritma yang di pakai adalah *naive bayes* di gunakan untuk memecahkan masalah tersebut, oleh karna itu penulis ingin menguji algoritma klasifikasi lain apakah tingkat akurasi nya lebih baik atau di bawah nilai dari *naive bayes*.

Data mining dapat digunakan oleh marketer untuk menganalisis pasar sehingga seorang market erdapat memahami pasar yang sudah ada ataupun menemukan peluang-peluang yang baru untuk meningkatkan keuntungan termasuk memprediksi kriteria target pasar (Han, Kamber, and Pei 2012). Oleh karena itu, penelitian ini fokus pada pemanfaatan *data mining* untuk memprediksi tingkat uji kelayakan *approval credit* calon nasabah kredit bank dari produk layanan perbankan, sehingga dapat diketahui apakah calon nasabah yang bersangkutan merupakan nasabah yang berpotensi menjadi nasabah kredit yang produktif atau tidak di liat dari penelitian sebelumnya algoritma yang di pakai adalah *Decision Tree* di gunakan untuk memecahkan masalah tersebut, oleh karna itu penulis ingin menguji algoritma klasifikasi lain apakah tingkat akurasi nya lebih baik atau di bawah nilai dari *decesion tree*.

2. TINJAUAN PUSTAKA

Tinjauan pustaka digunakan untuk mengetahui landasan dari judul yang diangkat dalam penelitian. Tinjauan pustaka merupakan bagian dari proposal penelitian dan sering kali menjadi sebuah bab tersendiri, untuk format penulisan dalam mensitasi sebuah penelitian ini menggunakan mendeley untuk style *American political science assosiation*

2.1 Neural Network

Neural network adalah satu set unit *input/output* yang terhubung dimana tiap relasinya memiliki bobot NN ini merupakan sistem adaptif yang dapat merubah strukturnya untuk memecahkan masalah berdasarkan informasi eksternal maupun internal yang mengalir melalui jaringan tersebut (Wu and Barnes 2011). Secara sederhana NN adalah sebuah alat pemodelan data statistik non-linear. NN dapat digunakan untuk memodelkan hubungan yang kompleks antara input dan output untuk menemukan pola-pola pada data. Neuron juga terdiri dari satu output. Outputnya adalah terbentuk dari pengolahan berbagai input oleh neuron (Kusrini et al. 2009).

2.2 Naive Bayes

Klasifikasi Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Klasifikasi Bayes juga dikenal dengan *Naïve Bayes*, *idiot's Bayes*, *simple Bayes*, dan *independence Bayes* (Wu and Barnes 2011). Klasifikasi Bayes didasarkan pada teorema Bayes, diambil dari nama seorang ahli matematika yang juga menteri Prebysterian Inggris, Thomas Bayes. *Naïve Bayes* merupakan sebuah model klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. *Naïve Bayes* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. Teknik *Naïve Bayes* (NB) adalah salah satu bentuk sederhana dari Bayesian yang jaringan untuk klasifikasi. Sebuah jaringan Bayes dapat dilihat sebagai diarahkan sebagai tabel dengan distribusi probabilitas gabungan lebih dari satu set diskrit dan variabel stokastik. Metode ini penting karena beberapa alasan, termasuk berikut. Hal ini sangat mudah untuk membangun, tidak perlu ada yang rumit. Parameter estimasi skema berulang. Ini berarti dapat segera diterapkan untuk besar data set. Sangat mudah untuk menafsirkan, sehingga pengguna tidak terampil dalam teknologi classifier dapat memahami mengapa itu adalah membuat klasifikasi itu membuat. Dan,

sangat penting, hal itu sering sangat baik: Ini mungkin bukan classifier terbaik dalam setiap diberikan aplikasi, tetapi biasanya dapat diandalkan untuk menjadi kuat dan melakukan dengan sangat baik (Wu and Barnes 2011).

2.3 K-Nearest Neighbor

Algoritma *k-nearest neighbor* (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut, Ketepatan algoritma k-NN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur, agar performa klasifikasi menjadi lebih baik, kNN juga merupakan contoh teknik *lazy learning*, yaitu teknik yang menunggu sampai pertanyaan (*query*) datang agar sama dengan data training (Wu and Barnes 2011).

Penelitian terdahulu yang terkait dengan topik aplikasi metode klasifikasi data mining pada analisa resiko kredit telah banyak dilakukan seperti penelitian yang dilakukan oleh Jing-xian Zhao tentang pembuatan model kombinasi neural network dan decision tree untuk resiko kredit, penelitian yang dilakukan oleh Eliana Angelini tentang Evaluasi resiko kredit dengan pendekatan neural network), Penelitian oleh Kin Keung Lai tentang analisa resiko kredit menggunakan reliability berbasis model *ensemble neural network* (Wu and Barnes 2011).

2.4 Pengumpulan Data

Penulis Memilih metode yang akan digunakan pada saat pengujian data. Metode yang dipilih, berdasarkan penelitian yang terdahulu. Penulis menggunakan Metode Algoritma *Neural Network*, *Naïve Bayes*, dan K-NN.

2.5. Evaluasi dan Validasi Hasil

2.5.1 Cross Validation

Cross validation adalah pengujian standar yang dilakukan untuk memprediksi *error rate*. Data *training* dibagi secara random ke dalam beberapa bagian dengan perbandingan yang sama kemudian *error rate* dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh *error rate* untuk mendapatkan *error rate* secara keseluruhan.

2.5.2 Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan 4 keluaran, yaitu: *recall*, *precision*, *accuracy* dan *error rate*. Evaluasi model klasifikasi didasarkan pada pengujian untuk memperkirakan obyek yang benar dan salah (Wu and Barnes 2011).

2.5.3 ROC Curve

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true positives* sebagai garis vertikal (Vercellis 2009).

Metode klasifikasi bisa dievaluasi berdasarkan kriteria seperti tingkat akurasi, kecepatan, kehandalan, skabilitas dan interpretabilitas (Vercellis 2009).

2.5.4 Validasi

Menurut (Florin Gorunescu 2011) Diperlukan cara yang sistematis untuk mengevaluasi kinerja suatu metoda. Evaluasi klasifikasi didasarkan pada pengujian pada obyek benar dan salah, menurut (Park et al. 2012) Validasi data digunakan untuk menentukan jenis terbaik dari skema belajar

yang digunakan, berdasarkan data pelatihan untuk melatih skema pembelajaran untuk memaksimalkan penggunaan data.

3. METODE PENELITIAN

Pengusaha di bidang jasa perbankan atau finace sangat menikmati fasilitas atau kemudahan yang diberikan oleh Pemerintah melalui kebijakan deregulasi tersebut. Ekspansi dan pembukaan kantor-kantor bank atau finace ini disamping memerlukan sejumlah tenaga kerja yang tidak sedikit, juga membutuhkan strategi pemasaran yang jitu dalam upaya menarik dana pihak ketiga untuk menyimpan uangnya di bank tersebut dan di pihak lain berusaha menyalurkan kredit yang disediakan ke pihak ketiga baik perorangan maupun perusahaan.

Data Mining didefinisikan sebagai proses penemuan pola dalam data. Menurut (Florin Gorunescu 2011) "Data mining adalah perpaduan dari ilmu statistik, kecerdasan buatan, dan penelitian bidang database". Nama data mining berasal dari kemiripan antara pencarian informasi yang bernilai dari database yang besar dengan menambang sebuah gunung untuk sesuatu yang bernilai. Keduanya memerlukan penyaringan melalui sejumlah besar material, atau menyelidiki dengan cerdas untuk mencari keberadaan sesuatu yang disebut bernilai tadi (Wu and Barnes 2011). Istilah lain dari data yaitu knowledge mining from databases, knowledge extraction, data/pattern analysis, data archeology, dan data dredging (Han and Gao 2010). Banyak yang menggunakan data mining sebagai istilah populer dari KDD.

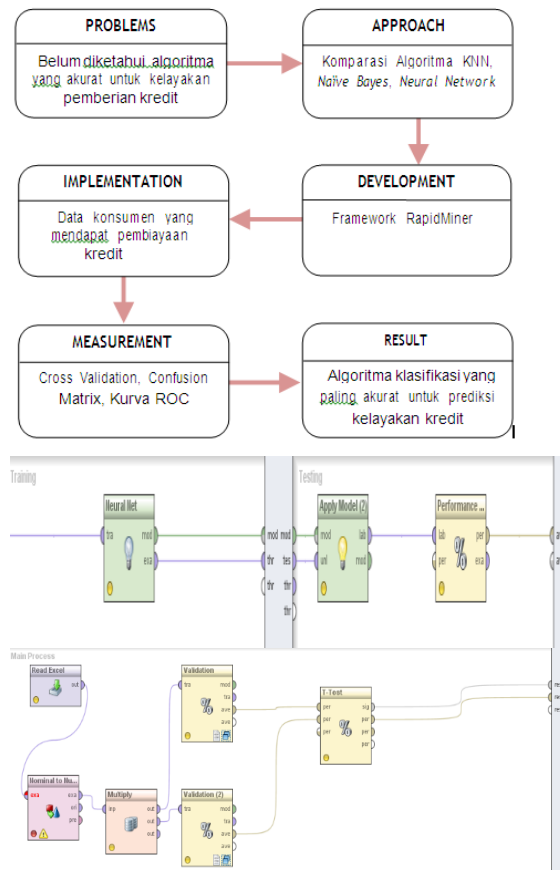
3.1 Kerangka Pemikiran

Dalam menyelesaikan penelitian, penulis membuat sebuah kerangka pemikiran yang berguna sebagai pedoman atau acuan penelitian ini sehingga penelitian dapat dilakukan secara konsisten.

Gambar 1. Kerangka pemikiran

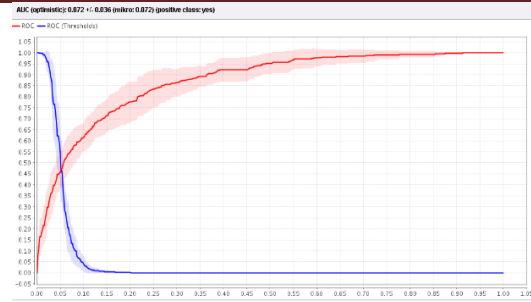
4. HASIL DAN PEMBAHASAN

Hasil dari pengujian model yang dilakukan adalah membandingkan algoritma mana yang lebih akurat dan memperbesar akurasi dengan menggunakan T-Test pada Algoritma pada *framework* RapidMiner dengan desain model berikut ini:



Gambar 2. Skema Pengujian Validation dan T-test pada

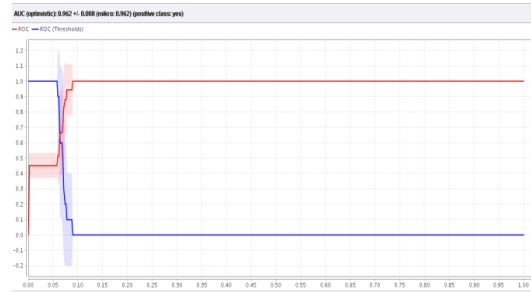
a. Hasil AUC Algoritma Neural Network



Gambar 3 Grafik AUC Neural Network

Kurva ROC yang dihasilkan berdasarkan pengujian data pada gambar di atas, menunjukkan bahwa ada peningkatan pada akurasi menggunakan Neural Network sebesar 89.81% dan AUC sebesar 0.882

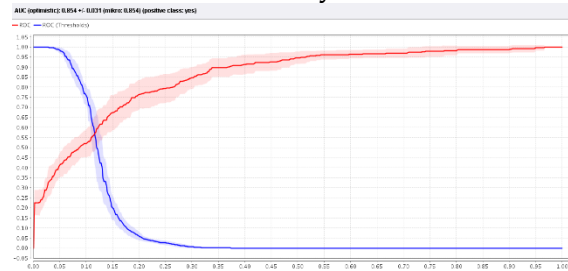
b. Hasil AUC K-NN



Gambar 4. Grafik AUC K-NN

Kurva ROC yang dihasilkan berdasarkan pengujian data pada gambar di atas, menunjukkan bahwa ada peningkatan pada akurasi menggunakan K-NN sebesar 84.70% dan AUC sebesar 0.962

c. Hasil AUC Naïve Bayes



Gambar 5. Grafik AUC Naïve Bayes

Kurva ROC yang dihasilkan berdasarkan pengujian data pada gambar di atas, menunjukkan bahwa ada peningkatan pada akurasi menggunakan Naïve Bayes sebesar 87.60% dan AUC sebesar 0.844

Tabel 1. Perbandingan Performace Algoritma

Algoritma	Accuracy	AUC
NN	89.81%	0.882
NB	84.60%	0.844
K-NN	87.79%	0.962

4.1 Pengujian dengan T-Test

Pengujian T-Test ini akan menguji algoritma Klasifikasi ini agar mendapatkan nilai yang terbaik, dimana dalam pengujian tersebut sampai mendapatkan nilai terkecil ≤ 0,05 dinyatakan sebagai hasil uji yang terbaik (Santoso. S:2010).

1. Hasil T-test antara algoritma KNN dengan Neural Network

T-Test Significance

	0.893 +/- 0.004	0.897 +/- 0.012
0.893 +/- 0.004		0.391
0.897 +/- 0.012		

4.2 Analisis Hasil Komparasi

Berdasarkan dari analisis pengujian masing-masing metode diatas maka dapat dirangkumkan hasilnya seperti Tabel berikut

**Tabel 2
Perbandingan Performace Algoritma Dengan T-Tes**

ALGORITMA	NB	NN	K-NN
NB	-	0.001	0.001
NN	0.000	-	0.009
K-NN	0.000	0.009	-

Melihat hasil perhitungan yang terangkum pada Tabel diatas dengan menerapkan klasifikasi performance keakurasian AUC maka diperoleh hasil penelitian yaitu, terdapat dua metode yang merupakan kategori *Good Clasification* yaitu untuk metode Neural Network dengan nilai AUC 0.993, K-NN dengan nilai AUC 0.962 dan metode algoritma NN dengan AUC 0.872 dan metode naïve bayes yang termasuk kategori *Fair Clasification* dengan nilai AUC

0.854.

Berdasarkan Tabel di atas juga dapat di lihat bahwa nilai akurasi untuk metode aloritma klasifikasi yang terbaik adalah Algoritma Neural Network memiliki Akurasi yang lebih tinggi dengan nilai 89.81% dibandingkan dengan 2 algoritma lain nya sedangkan di urutan ke dua algoritma K-NN dengan nilai 87.79% dan yang terakhir algoritma Naive Bayes dengan Nilai 84.70%.

5. KESIMPULAN

Dalam penelitian ini dilakukan pembuatan model menggunakan algoritma Klasifikasi yaitu *Neural Network*, Naive Bayes dan K-NN, menggunakan data pemasaran pada Bank. Algoritma *Neural Network* memiliki Akurasi yang lebih tinggi dengan nilai 89.81% dibandingkan dengan 2 algoritma lain nya sedangkan di urutan ke dua yaitu algoritma K-NN dengan nilai 87.79% dan yang terakhir algoritma *Naive Bayes* dengan Nilai 84.60%. Dengan demikian algoritma *Neural network* dapat memberikan pemecahan untuk permasalahan penentuan kelayakan konsumen yang mendapatkan pembiayaan kredit.

Pada kasus Prediksi uji kelayakan kredit *approval* pada produk layanan perbangkan menggunakan Algoritma Klasifikasi *data mining* dapat diterapkan pada data calon nasabah yang dihubungi untuk memprediksi keberhasilan uji kelayakan aproval kredit . Berdasarkan data set yang penelitian gunakan ini terbukti bahwa algoritma *Neural Network* ternyata lebih akurat bila dibandingkan dengan algoritma klasifikasi lain nya. Hal ini terlihat dari hasil evaluasi yang telah dilakukan. Dengan hasil ini, menunjukkan bahwa *Neural Network* merupakan metode yang cukup baik dalam prediksi data sehingga dapat memberikan hasil untuk permasalahan identifikasi calon nasabah.

Untuk keperluan penelitian lebih lanjut mengenai komparasi metode klasifikasi data *mining* dengan menggunakan data *additional* bank Ini maka disarankan untuk melakukan penyeleksian atribut, dikarenakan Atribut pday dalam metode algoritma tidak berpengaruh (hal ini di karenakan nilai nya sama) sehingga bisa dianalisa lebih lanjut

apakah atribut tersebut diperlukan atau tidak. Penelitian semacam ini dapat dikembangkan pada unit bisnis serupa atau yang lainnya. Penelitian ini dapat dikembangkan dengan algoritma yang lain misalkan saja dengan metode statistik lainnya seperti *Support Vector Machine*

REFERENSI

- Florin Gorunescu. 2011. *Data Maning : Concept and Techniques*. Verlag berlin Heidelberg: Springer.
- Han, Jiawei, and Jing Gao. 2010. "Research Challenges for Data Mining in Science and Engineering." : 1–18.
- Han, Jiawei, Micheline Kamber, and Jian Pei. 2012. *Data Mining Introduction*. <https://linkinghub.elsevier.com/retrieve/pii/B9780123814791000010>.
- Kusrini, Sri Hartati, Retantyo Wardoyo, and Agus Harjoko. 2009. "Perbandingan Metode Nearest Neighbor Dan Algoritma C4.5 Untuk Menganalisis Kemungkinan Pengunduran Diri Calon Mahasiswa Di STMIK AMIKOM Yogyakarta." *Jurnal Dasi* 10(1): 114–32.
- Park, Deuk Hee, Hyea Kyeong Kim, Il Young Choi, and Jae Kyeong Kim. 2012. "A Literature Review and Classification of Recommender Systems Research." *Expert Systems with Applications* 39(11): 10059–72. <http://dx.doi.org/10.1016/j.eswa.2012.02.038>.
- Vercellis, C. 2009. *Business Intelligent: Data Mining and Optimization for Decision Making*. Southern Gate: John Willey & Sons Inc.
- Wu, Chong, and David Barnes. 2011. "A Literature Review of Decision-Making Models and Approaches for Partner Selection in Agile Supply Chains." *Journal of Purchasing and Supply Management* 17(4): 256–74. <http://dx.doi.org/10.1016/j.pursup.2011.09.002>.