UNIVERSITAS GUNADARMA

PROGRAM STUDI DOKTOR TEKNOLOGI INFORMASI

PROPOSAL
SBK
(SEMINAR BIDANG KAJIAN)

TEMA

Pemodelan Data pada Proses *Waste to Energy (WTE)* untuk Menemukan Optimalisasi Kalori Bahan Baku Energi guna Mendukung Pencapaian Bauran Energi Baru Terbarukan (EBT)

Disusun oleh

NAMA : **LUQMAN**

NPM : **99220705**

TANGGAL : **31 Juli 2021**

Jakarta
2021

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam Rencana Usaha Penyediaan Tenaga Listrik (RUPTL) 2018-2027, penambahan pembangkit selama 10 tahun direncanakan total mencapai 56,024 gigawatt (GW). Penambahan pembangkit tersebut terdiri dari bauran energi antara lain batu bara 54,4%, energi baru terbarukan (EBT) 23%, gas 22,2% dan BBM 0,4% yang harus dicapai pada akhir tahun 2025.

Untuk memenuhi target dari EBT yang mempunyai porsi 23%, perlu dilakukan upaya yang komperhensif mulai dari sisi sumber energi, *suplay chain* dan sebuah model yang efesien agar EBT dapat memenuhi target sesuai RUPTL.

Teknologi informasi sangat berberan penting dalam upaya pencapaian target EBT yang telah ditetapkan oleh pemerintah, khususnya pengolahan data-data yang dihasilkan selama proses rantai penyediaan energi sehingga akan didapat sebuah model yang efesien dan optimal dalam pencapaian target EBT.

Berbagai bauran jenis EBT antara lain panas bumi, hydro, mini-micro hydro, bioenergi/ sampah *(Waste to Energy/ WTE)*, surya, angin dan gelombang laut dengan potensi di Indonesia seperti tabel 1 dibawah ini.

Table 1 Potensi Bauran EBT

No	Jenis Energi	Potensi	Kapasitas Terpasang	Pemanfaatan
1	Panas Bumi	29.544 MW	1.438,5 MW	4,9%
2	*Hydro*	75.091 MW	4.826,7 MW	6,4%
3	*Mini-micro Hydro*	19.385 MW	197,4 MW	1,0%
4	Bioenergi	32.654 MW	1.671,0 MW	5,1%
5	Surya	207.898 MW (4,80 kWh/m²/hari)	78,5 MW	0,04%
6	Angin	60.647 MW (≥ 4m/s)	3,1 MW	0,01%
7	Gelombang Laut	17.989 MW	0,3 MW	0,002%

Salah satu potensi bauran energi yang dapat dikembangkan adalah WTE. Potensi sampah yang besar dengan berbagai karakteristik dan varibel yang berbeda-beda menjadi tantangan untuk mendapatkan sebuah model yang optimal untuk memperoleh jenis pelet/breket dari sampah dengan kalori yang memenuhi standar untuk dijadikan sebagai bahan baku energi yang digunakan sebagai campuran batu bara / cofiring di PLTU, gasifier, kompor dan bolier mini atau sebagai bahan energi lain jika pelet/breket tersebut memiliki kalori diatas 3200 kcal/kg.

Upaya penelitian untuk mengolah sampah menjadi energi telah dilakukan dengan salah satunya adalah model Tempat Olah Sampah Setempat (TOSS) menggunakan mekanisme peuyemisasi, yaitu mengolah sampah tanpa memilah dan memberikan bioaktifator sebagai pemicu untuk proses mengurai sampah menjadi bahan baku pelet/ breket. Penelitian ini dilakukan dengan 2 (dua) sumber sampah yang berbeda, yaitu tipe sampah perkotaan untuk wilayah DKI Jakarta dan tipe sampah non perkotaan untuk wilayah Kabupaten Klungkung Bali.

Hasil dari penelitian diatas telah berhasil mendapatkan sebuah pelet/ breket dengan kalori yang berbeda-beda antara 1500 s/d 4000 kcal/kg, sehingga sangat sulit untuk pemanfaatan berikutnya dikarenakan tidak ada standar hasil produksi. Upaya mencari formulasi yang optimal telah dilakukan, tetapi karena banyaknya varibel yang mempengaruhi nilai kalori yang dihasilkan maka perlu dilakukan kajian lebih lanjut. Untuk mendapatkan sebuah produk dengan kalori yang memenuhi standar sangat dipengaruhi oleh banyak hal, antara lain adalah karakteristik sampah, komposisi bioaktifator, suhu, kelembapan dan lamanya waktu proses. Seluruh variabel yang mempengaruhi hasil kalori pelet/ breket **tidak** memiliki hubungan yang berbanding lurus, masing-masing mempunyai nilai optimal tersendiri, sehingga sangat sulit untuk menentukan standar maksimal dari kalori yang akan diperoleh.

Penelitian yang terkait dengan klusterisasi dan Named Entity Recognition (NER) yang mengolah kemiripan sumber data dan penelitian pengolahan sampah padat menjadi energi *(Waste to Energy)* telah dilakukan sampai saat ini sebagai bahan kajian awal antara lain :

1. *"Data analytics approach to create waste generation profiles for waste management and collection"*. Dalam tulisan ini, pendekatan berbasis data berbasis *self organizing maps* (SOM) dan algoritma *k-means* dikembangkan untuk membuat satu set profil jenis timbulan sampah. Pendekatannya adalah ditunjukkan menggunakan data pembobotan limbah tingkat kontainer ekstensif yang dikumpulkan di metropolitan area Helsinki, Finlandia. Hasil yang diperoleh menyoroti potensi analitik data tingkat lanjut pendekatan dalam menghasilkan informasi timbulan sampah yang lebih rinci misalnya untuk dasar layanan umpan balik yang disesuaikan bagi produsen limbah dan perencanaan serta optimalisasi pengumpulan dan daur ulang limbah. (Niska & Serkkola, 2018).
2. *"Agglomerative Fuzzy K-Means Clustering Algorithm with Selection of Number of Clusters"*. Dalam makalah ini, dijelaskan algoritma pengelompokan fuzzy K-

means aglomeratif untuk data numerik, perluasan dari algoritma fuzzy K-means standar dengan memperkenalkan istilah penalti ke fungsi tujuan untuk membuat proses pengelompokan tidak sensitif terhadap pusat cluster awal. Algoritma baru dapat menghasilkan pengelompokan yang lebih konsisten dari kumpulan pusat kluster awal yang berbeda. Dikombinasikan dengan teknik validasi cluster, algoritme baru dapat menentukan jumlah cluster dalam kumpulan data, yang merupakan masalah umum dalam clustering k-means. (Li et al., 2008) *(Jurnal IEEE, doi : 10.1109/TKDE.2008.88)*

3. *"Analysing efficiency of Waste to Energy Systems: Using Data Envelopment Analysis in Municipal Solid Waste Management"*, penelitian ini membahas model terintegrasi pengelolaan limbah padat, pengurangan limbah tepat di titik sumbernya sebelum memasuki rantai aliran limbah, penggunaan kembali limbah yang dihasilkan untuk pemulihan dengan cara daur ulang dan pembuangan melalui fasilitas pembakaran yang ramah lingkungan serta tempat pembuangan sampah yang memenuhi standar kebijakan seiring dengan perkembangannya. (Albores et al., 2016) *(Jurnal Elsevier, doi: 10.1016/j.proenv. 2016.07.007)*
4. *"Classification of groundwater chemistry in Shimabara, using self-organizing maps (SOM)"*, Penelitian ini dilakukan di kota Shimabara di Prefektur Nagasaki, Jepang, terletak di semenanjung vulkanik yang memiliki air tanah yang melimpah. Hampir semua pasokan air publik menggunakan air tanah di wilayah ini. Oleh karena itu, memahami karakteristik air tanah merupakan prasyarat untuk pengelolaan pasokan air yang tepat. Karenanya penentuan karakteristik kimia air tanah di Shimabara dengan menggunakan Self Organizing Map (SOM). (Nakagawa et al., 2017) *(doi :10.2166/nh.2016.072).*
5. *Jurnal* dengan judul *"A Bootstrapping Approach With CRF and Deep Learning Models for Improving the Biomedical Named Entity Recognition in Multi-Domains"* diterbitkan oleh IEEE tahun 2019 yang ditulis oleh Juae K, Youngjoong K dan Jungyun S dari Korea Selatan membahas sebuah metode yang secara otomatis menghasilkan *corpus* NER biomedis berlabel yang mencakup berbagai sub-domain dengan menggunakan kategori yang tepat dari kelompok semantik *Unified Medical Language System* (UMLS). Pembahasan menggunakan pendekatan *bootstrap* dengan sejumlah kecil *annotated corpus* yang secara otomatis akan menghasilkan sejumlah *corpus*

dan kemudian membangun sistem NER biomedis yang dilatih dengan *machine-labeled corpus*. (Kim et al., 2019)

6. *Jurnal dengan judul "Arabic named entity recognition using deep learning approach" diterbitkan oleh International Journal of Electrical and Computer Engineering (IJECE) tahun 2019 yang ditulis oleh Ismail El Bazi dan Nabil Laachfoubi dari IR2M Laboratory, FST, Univ Hassan 1st, Settat, Morocco* membahas NER berbahasa Arab yang sangat bergantung pada sumber daya eksternal dan rekayasa fitur untuk menghasilkan *state-of-the-art*. Untuk mengatasi keterbatasan tersebut, diusulkan untuk menggunakan pendekatan *deep learning* untuk menangani tugas NER bahasa Arab. Diperkenalkan juga *neural network architecture* dengan dua arah *Long Short-Term Memory* (LSTM) dan *Conditional Random Fields* (CRF) dan diujicobakan dengan berbagai *hyperparameter* yang umum digunakan untuk menilai pengaruhnya terhadap kinerja keseluruhan sistem. Model yang dibangun mendapatkan dua sumber informasi tentang kata-kata sebagai masukan yaitu kata yang telah dilatih sebelumnya dan representasi berbasis karakter, serta menghilangkan kebutuhan akan pengetahuan khusus atau rekayasa fiturnya. Hasilnya diperoleh *state-of-the-art* pada corpus ANERcorp dengan standar skor F1 90,6%. (El Bazi & Laachfoubi, 2019)
7. *Jurnal dengan judul "Effective Use of Bidirectional Language Modeling for Transfer Learning in Biomedical Named Entity Recognition" diterbitkan di Journal of Machine Learning Research tahun 2018 yang ditulis oleh Devendra S, Pengtao X, Mrinmaya S dan Eric P dari Machine Learning Department, CMU, Pittsburgh, PA, USA* membahas NER biomedis menjadi konsep dasar dalam penambangan teks dokumen medis dan memiliki banyak aplikasi. Pendekatan berbasis *deep leaning* untuk tugas ini telah mendapatkan perhatian yang meningkat dalam beberapa tahun terakhir karena parameternya dapat dipelajari *end to end* tanpa campur tangan manusia. Namun, pendekatan ini mengandalkan *high-quality labeled data* dan hal ini sangat mahal untuk diperoleh. Untuk mengatasi masalah tersebut dilakukan penelitian untuk mendapatkan cara menggunakan *unlabeled text data* untuk meningkatkan kinerja model NER. (Sachan et al., 2017)
8. *Jurnal dengan judul "Named-Entity Recognition for Indonesian Language using Bidirectional LSTM-CNNs" diterbitkan oleh Elsevier tahun 2018 yang ditulis*

oleh William Gunawan, Derwin Suhartono, Fredy Purnomo dan Andrew Ongko dari Indonesia menjelaskan implementasi NER untuk Bahasa Indonesia dengan menggunakan berbagai pendekatan deep learning, namun terutama difokuskan pada arsitektur hybrid Bidirectional LSTM (BLSTM) dan Convolutional Neural Network (CNN). Sudah ada beberapa NER untuk pengembangan beberapa bahasa, antara lain bahasa Inggris, Vietnam, Jerman, India dan banyak lagi lainnya. Namun penelitian pada jurnal ini berfokus pada bahasa Indonesia. NER bahasa Indonesia ini berhasil mengekstrak informasi dari artikel ke dalam 4 kelas yang berbeda, yaitu Orang, Organisasi, Lokasi, dan Kejadian. (Gunawan et al., 2018)

9. *Jurnal dengan judul "An Analysis of the Performance of Named Entity Recognition over OCRed Documents" diterbitkan oleh Journal European Union's Horizon Research pada tahun 2020 yang ditulis oleh Ahmed H, Axel J, Nicolas S, Mickaël C, Antoine D dari Prancis* dibahas terkait penggunaan perpustakaan digital yang membutuhkan kemudahan akses ke dokumen yang sangat dipengaruhi oleh kualitas pengindeksan dokumen. Named Entity adalah salah satu informasi terpenting untuk mengindeks dokumen digital. Menurut studi terbaru, 80% dari 500 query teratas yang dikirim ke portal perpustakaan digital berisi setidaknya satu Named Entity. Namun sebagian besar dokumen digital diindeks melalui versi OCRed yang terjadi banyak kesalahan sehingga dapat menghalangi akses ke dokumen tersebut. (Filannino et al., 2013)

Dari penelaahan beberapa jurnal diatas diharapkan menjadi acuan awal untuk merumuskan tahapan penelitian dan sebagai bahan kajian awal sampai dimana pemanfaatan metode yang telah digunakan untuk tipikal data yang memiliki karakteristik yang sama dengan rencana data yang akan digunakan dalam penelitian desertasi ini.

Terkait dengan upaya untuk mendapatkan hasil kalori yang sesuai standar, maka perlu dilakukan pengumpulan data seluruh variabel sebagai data set, kemudian dilakukan analisa terhadap data set tersebut dengan menggunakan ***Self Organizing Maps (SOM)*** *dan* ***Fuzzy K-Means Clustering*** untuk kemudian dirumuskan sebuah model baru yang optimal dari keterhubungan seluruh data-data tersebut.

Ditinjau dari uraian latar belakang masalah, maka penulis mengambil sebuah tema desertasi *"**Pemodelan Data pada Proses Waste to Energy (WTE) untuk Menemukan Optimalisasi Kalori Bahan Baku Energi guna Mendukung Pencapaian Bauran Energi Baru Terbarukan (EBT)"***

Diharapkan dengan temuan model baru terkait hubungan setiap data/ variabel, maka akan menentukan upaya selanjutnya untuk mengoptimalisasi potensi bauran energi dari sampah mulai dari proses pengumpulan sampah, standar sensor pada peralatan sampai dengan rantai pasok pendistribusian sumber energi di seluruh Indonesia.

1.2 Rumusan Masalah Penelitian

Sesuai latar berlakang masalah diatas, dapat diambil rumusan masalah penelitian sebagai berikut :

1. Bagaimana karakteriktik data pengujian yang digunakan untuk menghasilkan model proses WTE yang optimal ?
2. Bagamana penerapan metode klaterisasi untuk menemukan model baru optimalisasi kalori produk WTE ?
3. Bagamana penerapan lanjutan temuan model pengelolaan WTE agar optimal ?

1.3 Tujuan Penelitian

Tujuan dari penelitian pengelolaan Waste to Energy dengan model baru tersebut adalah :

1. Mendapatkan sebuah sistem *clastering* yang dapat mengolah data karakteristik sampah dan karakteristik proses yang optimal.
2. Menemukan sebuah model pengelolaan sampah menjadi energi dengan karakteristik masing-masing yang disesuaikan dengan lokasi sumber sampah dengan hasil kalori optimal.

1.4 Kontribusi Penelitian

Hasil dari penelitian ini adalah sebuah model awal pengelolaan sampah menjadi energi dengan menggunakan data karakteriktik sampah yang diharapkan dapat berkontribusi sebagai berikut :

1. Dapat menetapkan sebuah Standar Nasional Indonesi (SNI) dalam pengelolaan sampah menjadi pellet sebagai bahan baku energi.

a. Standar proses/ pengelolaan
b. Standar peralatan/ mesin
c. Standar sensor

2. Dapat mengembangkan pasokan pellet yang stabil dengan diketahuinya potensi energi sampah yang optimal disetiap wilayah.
3. Dapat mengembangkan rantai tata niaga pellet sehingga dapat diketahui pemenuhan kebutuhan sumber pellet sebagai bahan baku energi.
4. Dapat dikembangkan sesuah sistem tata niaga pellet dengan diketahuinya petensi disetiap wilayah.
5. Menjadi dasar dalam menetapkan sebuah kebijakan pengelolaan energi di Indonesia.

BAB II
STUDI LITERATUR

2.1 Landasan Teori

Landasan teori berisi mengenai teori – teori yang berkaitan dengan dengan penelitian yang sedang dilakukan.

2.1.1 Energi Terbarukan (Renewable Energi)

International Energy Agency (IEA) mendefinisikan, energi terbarukan (renewable energy) merupakan energi yang berasal dari proses alam dan mengalami siklus berkelanjutan. IEA mengelompokan energi terbarukan seperti angin, biomassa, biofuel, hydrogen, hydropower, matahari, laut dan panas bumi (Asriyati, 2019). Pada tahun 1970 -an merupakan dimulainya konsep energi terbarukan, dengan tujuan untuk mengembangkan bahan bakar fosil dan nuklir. Secara universal, energi terbarukan adalah energi yang mudah “pulih kembali” oleh proses alam dengan proses siklus yang terus – menerus, energi terbarukan merupakan energi *sustainable*, artinya tersedia dalam waktu jauh ke depan.

2.1.2 Clustering

Definisi *clustering* adalah proses untuk menghimpun data ke dalam sejumlah kelompok sehingga data yang berada dalam satu klaster mempunyai tingkat kesamaan yang maksimum dan data antar klaster mempunyai kesamaan yang rendah (Zhang, Huang, & Tan, 2006).

Definisi lain metode clustering adalah pengelompokan didasarkan ukuran kecendrungan (kedekatan) dengan mempertimbangkan pendekatan untuk mencari kemiripan dalam data dan menyimpan data yang sama dalam suatu kelompok. Metode clustering memetakan himpunan data ke sejumlah kelompok, yang artinya kemiripan dalam suatu kelompok merupakan lebih besar diantara kelompok yang lainnya (Xu & Wunsch, 2009).

Konsep metode clustering adalah mengoptimumkan pusat cluster (Kusumadewi & Purnomo, Aplikasi Logika Fuzzy Edisi 2, 2010). Beberapa metode clustering yang umum digunakan :

a. Bebasis metode statistik seperti *Hirarchical Clustering Method* dan *Non Hirarchical Clustering Method*.

b. Berbasis Fuzzy : Fuzzy C – Means (FCM).
c. Berbasis Neural Network : Kohonen *Self-Organizing Maps* (SOM), *Learning Vector Quantization* (LVQ)

2.1.3 Kohonen *Self Organizing Maps (SOM)*

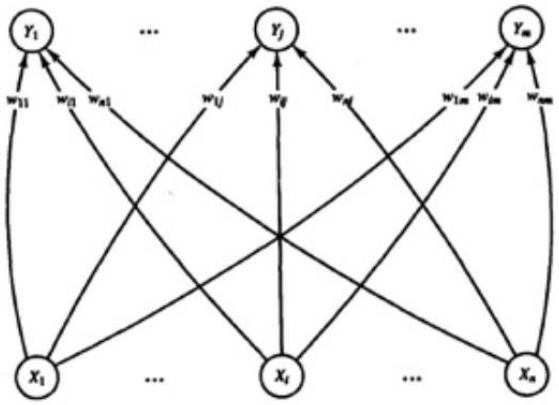
Self-Organizing Map (SOM) atau sering disebut topology-preserving map pertama kali diperkenalkan oleh Teuvo Kohonen pada tahun 1996. SOM merupakan salah satu teknik dalam Neural Network yang bertujuan untuk melakukan visualisasi data dengan cara mengurangi dimensi data melalui penggunaan self-organizing neural networks sehingga manusia dapat mengerti high-dimensional data yang dipetakan dalam bentuk low-dimensional data. Metode pembelajaran yang digunakan SOM adalah tanpa bimbingan dari suatu data input-target atau unsupervised learning yang mengasumsikan sebuah topologi yang terstruktur menjadian unit-unit kelas/cluster (Kohonen, 1989 dan Fausett, 1993).

Pada algoritma SOM, vektor bobot untuk setiap unit cluster berfungsi sebagai contoh dari input pola yang terkait dengan cluster itu. Selama proses self-organizing, cluster satuan yang bobotnya sesuai dengan pola vektor input yang paling dekat (biasanya, kuadrat dari jarak Euclidean minimum) dipilih sebagai pemenang. Unit pemenang dan unit tetangganya (dalam pengertian topologi dari unit cluster) terus memperbarui bobot merek (Fausett, 1993). Setiap output akan bereaksi terhadap pola input tertentu sehingga hasil Kohonen SOM akan menunjukkan adanya kesamaan ciri antar anggota dalam cluster yang sama.

Dalam jaringan SOM, neuron target tidak diletakkan dalam sebuah baris seperti layaknya model JST yang lain. Neuron target diletakkan dalam dua dimensi yang bentuk/topologinya dapat diatur. Topologi yang berbeda akan menghasilkan neuron sekitar neuron pemenang yang berbeda sehingga bobot yang dihaslkan juga akan berbeda. Pada SOM, perubahan bobot tidak hanya dilakukan pada bobot garis yang terhubung ke neuron pemenang saja, tetapi juga pada bobot garis ke neuron-neuron di sekitarnya. neuron di sekitar neuron pemenang ditentukan berdasarkan jaraknya dari neuron pemenang.

2.1.3.1 Arsitektur Topologi SOM

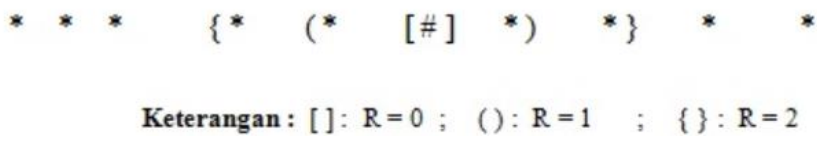
Arsitektur SOM merupakan jaringan yang terdiri dari dua lapisan (layer), yaitu lapisan input dan lapisan output. Setiap neuron dalam lapisan input terhubung dengan setiap neuron pada lapisan output. Setiap neuron dalam lapisan output merepresentasikan kelas (cluster)dari input yang diberikan.



Gambar 1. Arsitektur JST SOM (Fausett, 1993)

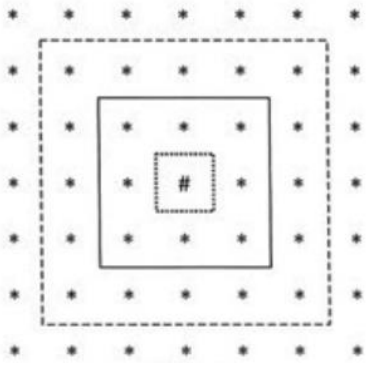
Sedangkan untuk topologi, SOM memiliki 3 jenis topologi hubungan ketetanggaan (neighborhood) yaitu linear array, rectangular dan heksagonal grid.

Topologi linear aray menunjukkan cluster unit yang tersusun secara linear. Cluster unit yang menjadi pemenang [#] memiliki dua unit tetangga (neighbour) yang berjarak 1 (R = 1), dan mempunyai dua unit tetangga yang berjarak 2 (R = 2).



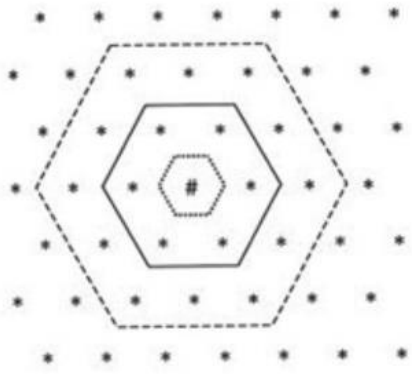
Gambar 2. Linier Array Unit (Fausett, 1993)

Rectangular grid adalah topologi dari cluster unit dua dimensi. Unit tetangga (neighbour) dari unit pemenang membentuk bujur sangkar. Unit pemenang [#] memiliki 8 neighbour berjarak 1 (R=1) dan 16 neighbour berjarak 2 (R=2).



Gambar 3. Rectangular Grid

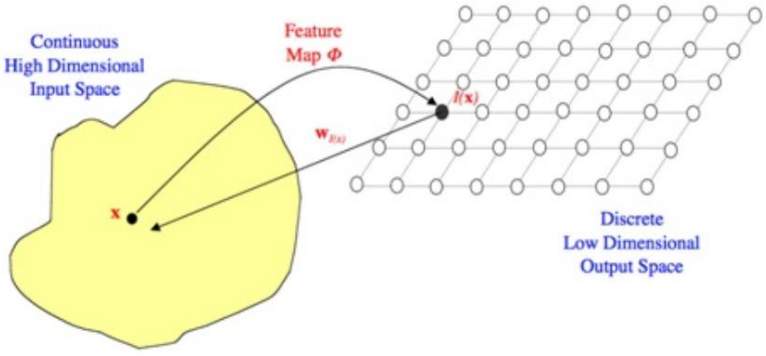
Dalam topologi heksagonal grid, unit tetangga (neighbour) yang berjarak 1 (R=1) dari unit pemenang adalah 6 dan yang berjarak 2 (R=2) adalah 12.



Gambar 4. Heksagonal Grid

1. Cara Kerja SOM

Secara umum cara kerja SOM dapat dilihat pada gambar dibawah ini :



Gambar 5. Cara Kerja SOM secara umum

Terdapat titik (x) pada ruang input untuk dipetakan ke titik $I_{(x)}$ pada ruang output. Setiap titik (I) dalam ruang output akan memetakan ke titik yang sesuai dalam ruang input melalui bobot $w_{I(x)}$.

Menurut Haykin (1999) terdapat tiga komponen penting dalam SOM yaitu :

a. Competition: Untuk setiap pola input, neuron menghitung nilai masing-masing fungsi diskriminan yang memberi dasar untuk kompetisi. Neuron tertentu dengan nilai terkecil dari fungsi diskriminan dinyatakan sebagai pemenang.
b. Cooperation: Neuron pemenang menentukan lokasi spasial dari lingkungan topologi excited neuron untuk memberi dasar kerjasama dalam suatu lingkungan neuron.
c. Synaptic Adaption: Excited neuron menurunkan nilai fungsi diskriminan yang berkaitan dengan pola input melalui penyesuaian bobot terkait sehingga respon dari neuron pemenang keaplikasi berikutnya dengan pola input yang sama akan meningkat.

Pada prinsipnya algoritma SOM mempunyai 2 proses perhitungan matematika, yaitu pada proses pencarian nilai bobot yang sesuai dengan nilai masukan dan perubahan nilai bobot yang telah ditemukan dengan jarak terdekat.

Perhitungan perubahan nilai bobot sekitar atau tetangga bobot pemenang tidak dihitung atau diberi nilai 0. Pemberian nilai ini dimaksudkan agar tiap bobot diarahkan ke nilai masukan sehingga nilai bobot akan mendekati nilai masukan.

Berikut merupaka algoritma SOM :

1. Neuron pada lapisan input (neuron input) sebanyak n dinotasikan sebagai x1, x2, ..., xn dan neuron pada lapisan output (neuron output) sebanyak m dinotasikan sebagai y1, y2, ..., ym. Bobot koneksi antara neuron input dan output dinotasikan sebagai wijditentukan secara acak antara 0 dan 1.
2. Selama kondisi penghentian bernilai salah, lakukan langkah 3 – 8.

3. Untuk setiap masukan (x1, x2, ..., xn) lakukan langkah 4 – 6.
4. Hitung jarak vektor input terhadap bobot koneksi dj untuk masing-masing neuron output dengan menggunakan rumus :

$$d_j = \sum_{i=1}^{n} (w_{ij} - x_i)^2$$

5. Cari indeks j di mana d_j minimum.
6. Untuk setiap w_{ij}, perbarui bobot koneksi dengan menggunakan rumus : $w_{ij}(t+1) = w_{ij}(t) + y(t)h_{ib}(t)\,(x_i(t) - w_{ij}(t))$
7. Modifikasi laju pemahaman
8. Uji kondisi penghentian

2.1.4 Fuzzy K-Means Clustering

Menurut Witten dan Frank (2005:137), k-means merupakan teknik clustering klasik. Pertama-tama perlu ditentukan jumlah cluster, ini merupakan parameter k. Kemudian poin k ditentukan secara acak sebagai pusat dari cluster. Semua instance ditetapkan ke pusat cluster terdekat berdasarkan perhitungan metrik jarak Euclidean. Selanjutnya centroid atau mean dari instances pada setiap cluster dihitung – ini merupakan bagian "means". Centroids ini kemudian menjadi nilai center baru untuk masing-masing cluster. Lalu proses di atas diulang sampai semua instance telah ditetapkan ke dalam cluster-cluster dan nilai center dari cluster tidak berubah lagi (sudah stabil).

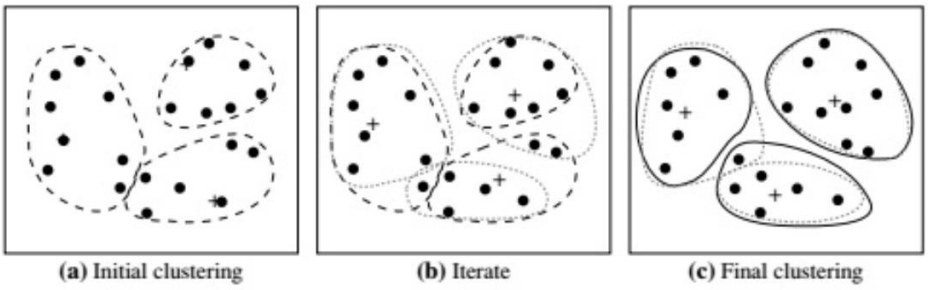
Pada sebuah artikel di situs datascience, Trevino (2016) menerangkan bahwa K-means clustering merupakan sebuah tipe unsupervised learning, dimana teknik ini digunakan ketika data yang ada belum mempunyai label/unlabeled data. Tujuan dari algoritma ini adalah untuk menemukan grup-grup data, dengan jumlah grup data direpresentasikan oleh variabel k. Algoritma ini bekerja secara iteratif untuk menempatkan setiap poin data ke dalam sebuah grup dari sejumlah k grup berdasarkan fitur yang disediakan. Poin data dikelompokkan berdasarkan fitur similaritas. Adapun hasil dari algoritma k-means clustering adalah :

1. Centroid dari cluster-cluster yang telah dibentuk, dimana centroid tersebut dapat digunakan untuk melabeli data baru.
2. Label untuk training data (setiap poin data ditempatkan pada sebuah cluster)

Pada k-means, biasanya diberikan sebuah set data, D, dari sejumlah n objek, dan k, jumlah cluster yang akan dibentuk. Algoritma partitioning mengorganisasikan objek ke dalam partisi/cluster ($k \leq n$). Dari cluster-cluster yang telah terbentuk, dapat dihitung jaraknya untuk mengetahui objektivitas dari ketidaksamaan antar cluster (Han, Kamber, dan Pei, 2012:451).

Berikut adalah penjelasan singkat mengenai proses clusteringdengan menggunakan metode k-means :

1. Tentukan jumlah cluster (k) yang diinginkan.
2. Tentukan nilai mean yang akan menjadi pusat cluster awal.
3. Tetapkan setiap objek ke dalam cluster berdasarkan nilai mean objek yang paling mirip.
4. Perbarui nilai mean dari cluster, yaitu dengan menghitung nilai meanobjek untuk setiap cluster.
5. Ulangi langkah 2-4 sampai tidak ada lagi perubahan pada nilai mean dari cluster. Ilustrasi dapat dilihat pada gambar 6.



Gambar 6. Ilustrasi proses k-means

Lebih lanjut Trevino (2016) juga menjabarkan kegunaan k-means clustering pada bisnis, dimana teknik ini bisa digunakan untuk mengkonfirmasi asumsi bisnis tentang tipe grup yang ada saat ini atau untuk mengidentifikasi grup-grup yang tidak diketahui pada data sets yang kompleks. Saat algoritma sudah dijalankan dan grup sudah didefinisikan, data baru dapat dengan mudah ditempatkan ke dalam grup yang tepat.

Contoh kegunaan dari k-means clustering adalah :

- Behavorial segmentation :

    - Segment by purchase history
    - Segment by activities on application, website, or platform
    - Define personas based on interests
    - Create profiles based on activity monitoring
- Inventory categorization :
    - Group inventory by sales activity
    - Group inventory by manufacturing metrics
- Sorting sensor measurements :
    - Detect activity types in motion sensors
    - Group images
    - Separate audio
    - Identify groups in health monitoring
- Detecting bots or anomalies :
    - Separate valid activity groups from bots
    - Group valid activity to clean up outlier detection

2.1.5 Named Entity Recognition (NER)

Dokumen adalah suatu media yang dapat memiliki informasi yang berarti, dokumen tersebut dapat berupa gambar dan teks. Dokumen teks adalah dokumen yang berisikan kumpulan dari karakter-karakter yang mejadi suatu kalimat. Jika pada dokumen bentuk teks biasanya teks tersebut memiliki informasi yang sangat penting, diantaraya berupa orang, nama, organisasi dan nama tempat. Cara untuk mendapatkan informasi dalam dokumen bentuk teks masih harus dilakukan secara konvensional, yaitu dengan cara membaca terlebih dahulu untuk semua dokumen tersebut, lalu dilanjtukan dengan menentukan kata atau kalimat yang mengandung karakteristik atau unik sesuai dengan yang ditentukan dan juga dengan melakukan hal tersebut, pastinya akan memakan waktu yang lebih lama dalam melakukan penetuannya untuk mendapatkan informasi dalam dokumen teks tersebut. Oleh sebab itu, perlu dibuatkannya NER atau *Named Entity Recognition* yang memiliki fungsi sebagai sesuatu yang dilakukan untuk mendapatkan informasi dari suatu dokumen yang bersifat penting, yang biasanya meliputi orang, nama, nama tempat atau informasi dari kumpulan data. Oleh karena itu dengan digunakannya konsep

NER atau Named Entity Recognition maka untuk mendapatkan informasi penting bisa dilakukan dengan cepat dan akurat.

NER merupakan bagian dari proses text mining dan natural language processing yang digunakan pada proses ekstraksi informasi. Tugas utama dari metode NER adalah untuk mengidentifikasi dan mengklasifikasikan nama dalam teks ke dalam kelas-kelas yang telah ditentukan.

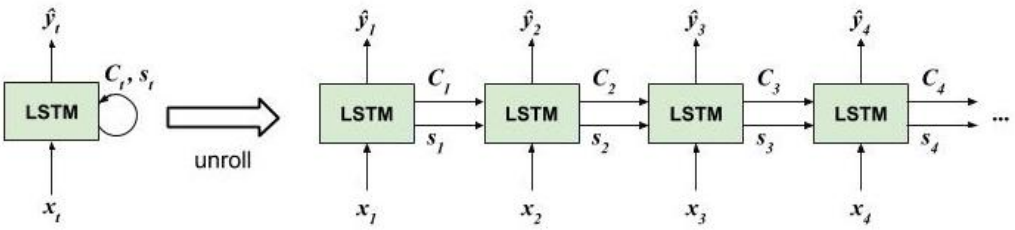
2.1.5.1 Conditional Random Field

Conditional Random Field merupakan sebuah model grafik yang digunakan untuk menghitung probabilitas nilai bersyarat dengan node output yang dihasilkan untuk digunakan sebagai input pada node lain. Conditional Random Field (CRF) merupakan metode campuran antara metode Maximum Entropy Markov Model (MEMM) dengan metode Hidden Markov Model (HMM). Terdapat 3 (tiga) tahapan pemodelan Conditonal Random Field, yaitu pembentukan model, inferensi dan decoding.

2.1.5.2 Long Short Term Memory (LSTM)

LSTM adalah salah satu jenis Recurrent Neural Network (RNN) dimana dilakukan sebuah modifikasi pada RNN dengan menambahkan cell memori yang dapat menyimpan informasi dalam jangka waktu yang lama. LSTM sebagai solusi jika terjadi vanishing gradient pada RNN saat memproses data sequential yang cukup panjang.

LSTM (Long Short Term Memory) adalah jenis modul pemrosesan lain pada RNN. LSTM temukan oleh Schmidhuber dan Hochreiter pada tahun 1997 dan kemudian dikembangkan dan dipopulerkan oleh banyak peneliti. Seperti hal nya RNN, jaringan LSTM (LSTM network) juga terdiri dari modul-modul yang melakukan pemrosesan berulang. Perbedaanya adalah modul-modul yang membentuk jaringan LSTM adalah modul LSTM.

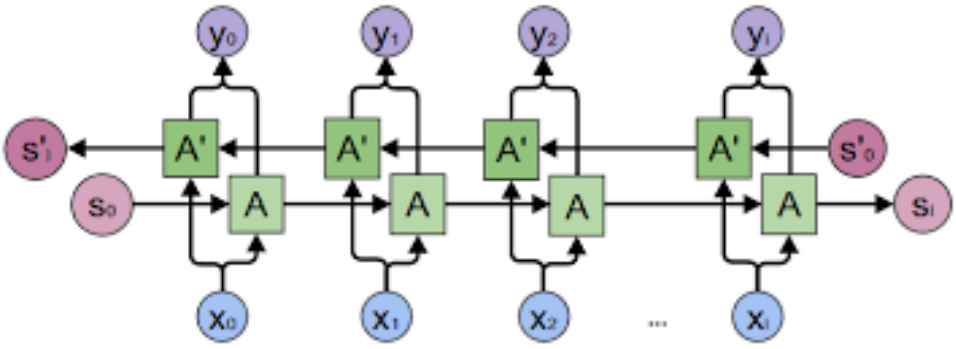


Gambar 7. Jaringan LSTM

Modul LSTM (kotak warna hijau) mempunyai pemrosesan yang berbeda dengan modul RNN biasa. Perbedaan lain adalah adanya tambahan sinyal yang diberikan dari satu langkah waktu ke langkah waktu selanjutnya, yaitu konteks, direpresentasikan dengan simbol Ct.

2.1.5.3 Bidirectional LSTM (BiLSTM)

Bidirectional LSTM adalah salah satu varian dari LSTM yang umum digunakan. Input yang dimasukkan ke dalam BiLSTM ada 2 jenis yaitu input *forward* dan input *backward*. Output dari lapisan ini umumnya digabungkan menjadi satu. Dengan layer ini, model dapat mempelajari informasi masa lalu (past) dan informasi masa mendatang (future) untuk tiap sekuen input.

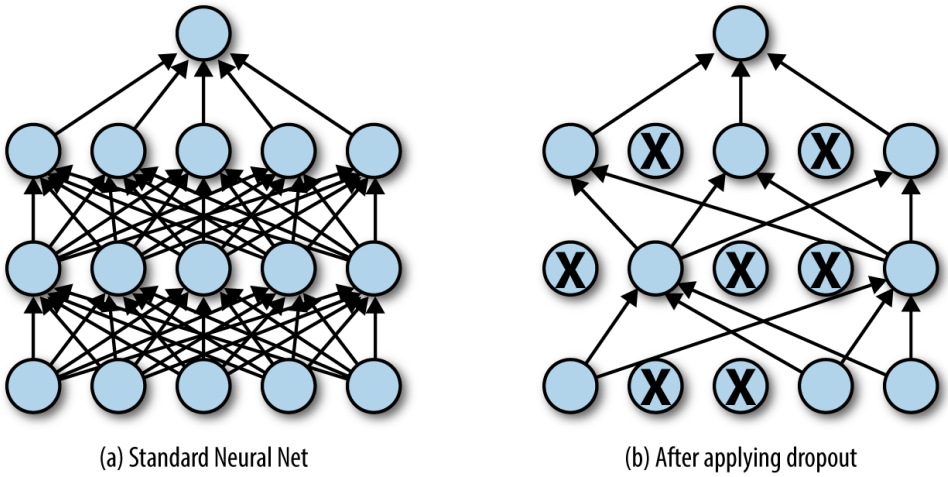


Gambar 8. Arsitektur BiLSTM

2.1.5.4 *Dropout*

Dropout merupakan teknik regularisasi model jaringan syaraf tiruan untuk mengurangi overfitting pada dataset. Dropout merupakan cara yang efisien untuk menghindari *overfitting*. Deep neural network pada umumnya memiliki lapisan yang banyak yang memungkinkan sebuah model mempelajari relasi kompleks antara input dan output. Dengan jumlah dataset yang terbatas, model deep neural network mungkin justru akan mempelajari relasi kompleks noise dari dataset. Dropout dilakukan

dengan cara mematikan (dropping out) unit dari lapisan tersembunyi (hidden layer) maupun lapisan yang tampak (visible layer).



Gambar 9. Regulasi Dropout

2.2 Kajian Penelitian

Sebagai upaya untuk melihat perkembangan penelitian dan kajiannya terkait dengan tema dan metoda yang akan digunakan dalam desertasi/ penelitian ini berikut beberapa kajian yang dipakai sebagai landasan :

Kajian pertama diambil dari jurnal *Elsevier* tahun 2018 dengan judul *"Waste Management",* dengan judul *"Data analytics approach to create waste generation profiles for waste management and collection"*. Dalam tulisan ini, pendekatan berbasis data berbasis *self organizing maps* (SOM) dan algoritma *k-means* dikembangkan untuk membuat satu set profil jenis timbulan sampah. Pendekatannya adalah ditunjukkan menggunakan data pembobotan limbah tingkat kontainer ekstensif yang dikumpulkan di metropolitan area Helsinki, Finlandia. Hasil yang diperoleh menyoroti potensi analitik data tingkat lanjut pendekatan dalam menghasilkan informasi timbulan sampah yang lebih rinci misalnya untuk dasar layanan umpan balik yang disesuaikan bagi produsen limbah dan perencanaan serta optimalisasi pengumpulan dan daur ulang limbah.

Kajian kedua diangkat dari jurnal IEEE tahun 2006 dengan judul *"Agglomerative Fuzzy K-Means Clustering Algorithm with Selection of Number of Clusters"*. Dalam jurnal tersebut dijelaskan algoritma pengelompokan fuzzy K-means aglomeratif untuk data numerik, perluasan dari algoritma fuzzy K-means standar dengan

memperkenalkan istilah penalti ke fungsi tujuan untuk membuat proses pengelompokan tidak sensitif terhadap pusat cluster awal. Algoritma baru dapat menghasilkan pengelompokan yang lebih konsisten dari kumpulan pusat kluster awal yang berbeda. Dikombinasikan dengan teknik validasi cluster, algoritma baru dapat menentukan jumlah cluster dalam kumpulan data, yang merupakan masalah umum dalam clustering k-means.

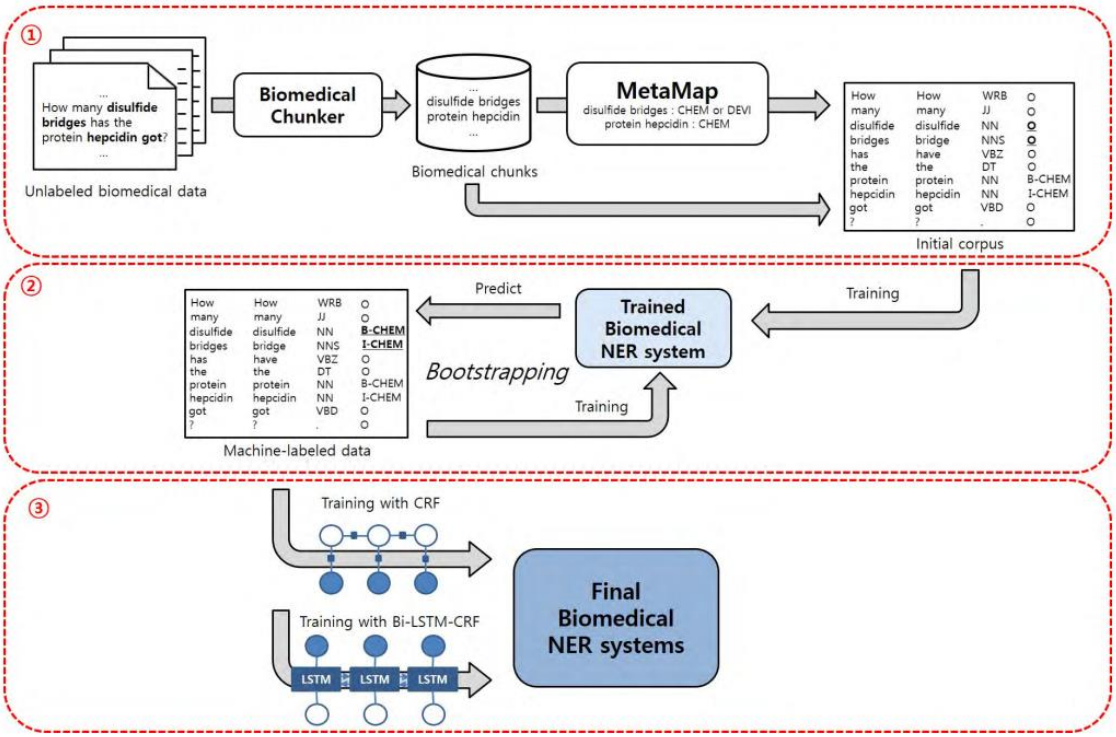
Kajian ketiga diambil dari jurnal *Elsevier* tahun 2016 dengan judul *"Analysing efficiency of Waste to Energy Systems: Using Data Envelopment Analysis in Municipal Solid Waste Management"*, penelitian ini membahas model terintegrasi pengelolaan limbah padat, pengurangan limbah tepat di titik sumbernya sebelum memasuki rantai aliran limbah, penggunaan kembali limbah yang dihasilkan untuk pemulihan dengan cara daur ulang dan pembuangan melalui fasilitas pembakaran yang ramah lingkungan serta tempat pembuangan sampah yang memenuhi standar kebijakan seiring dengan perkembangannya.

Kajian keempat diambil *Hydrology Research* yang diterbitkan oleh IWA Publishing Jepang tahun 2016 dengan judul *"Classification of groundwater chemistry in Shimabara, using self-organizing maps (SOM)"*, Penelitian ini dilakukan di kota Shimabara di Prefektur Nagasaki, Jepang, terletak di semenanjung vulkanik yang memiliki air tanah yang melimpah. Hampir semua pasokan air publik menggunakan air tanah di wilayah ini. Oleh karena itu, memahami karakteristik air tanah merupakan prasyarat untuk pengelolaan pasokan air yang tepat. Karenanya penentuan karakteristik kimia air tanah di Shimabara dengan menggunakan Self Organizing Map (SOM).

Dalam pembahasan pada jurnal yang kelima dengan judul "*A Bootstrapping Approach With CRF and Deep Learning Models for Improving the Biomedical Named Entity Recognition in Multi-Domain*s" (IEEE, 2019), dibahas masalah *Named Entity Recognition* biomedis (NER biomedis) yang merupakan komponen inti untuk membangun sistem pemrosesan teks biomedis, seperti sistem temu kembali informasi biomedis dan sistem penjawab pertanyaan. Saat ini banyak penelitian *machine learning* telah dikembangkan untuk NER biomedis. Pendekatan berbasis *machine learning* umumnya membutuhkan jumlah yang banyak dari *annotated corpus* untuk mencapai kinerja yang tinggi. Selain itu, sebagian besar *corpus* yang ada berfokus pada beberapa sub-domain tertentu, seperti penyakit, protein, dan spesies. Sulit bagi

sistem NER biomedis yang dilatih dengan *corpus* ini untuk memberikan banyak informasi untuk sistem pemrosesan teks biomedis.

Pada jurnal yang kelima ini diusulkan sebuah metode yang secara otomatis menghasilkan *corpus* NER biomedis berlabel yang mencakup berbagai sub-domain dengan menggunakan kategori yang tepat dari kelompok semantik *Unified Medical Language System* (UMLS). Pembahasan menggunakan pendekatan *bootstrap* dengan sejumlah kecil *annotated corpus* yang secara otomatis akan menghasilkan sejumlah *corpus* dan kemudian membangun sistem NER biomedis yang dilatih dengan *machine-labeled corpus*.



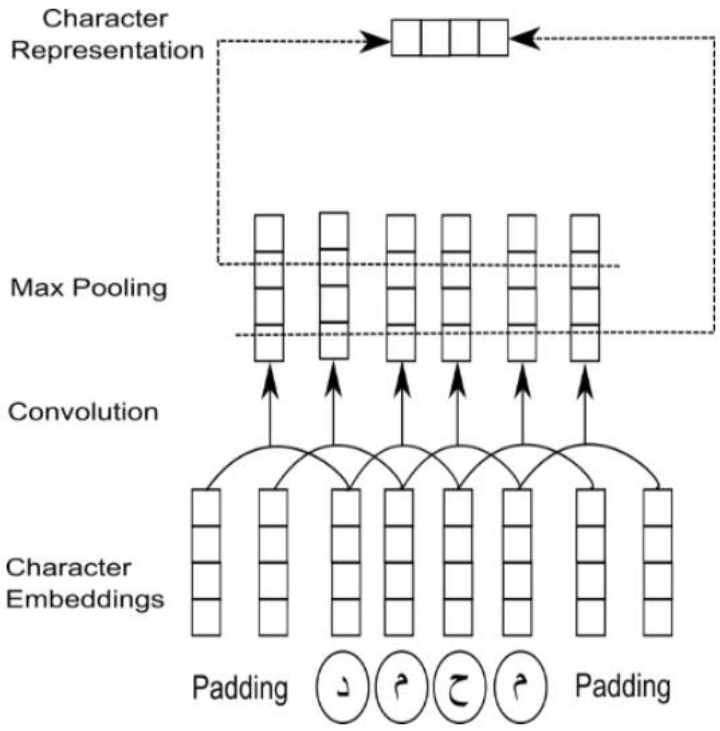
Gambar 10. Gambaran Umum metode yang diusulkan

Terakhir dilakukan percobaan dengan melatih dua *machine learning-based classifiers*, *conditional random fields* (CRF) dan *long short-term memory* (LSTM), dengan *machine-labeled data* untuk meningkatkan kinerja seperti gambar 10. Hasil percobaan menunjukkan bahwa metode yang diusulkan efektif untuk meningkatkan kinerja tersebut. Hasilnya, model yang diusulkan memperoleh kinerja yang lebih tinggi dalam 23,69% dibandingkan model yang hanya melatih sejumlah kecil *annotated corpus* secara manual dalam skor F1.

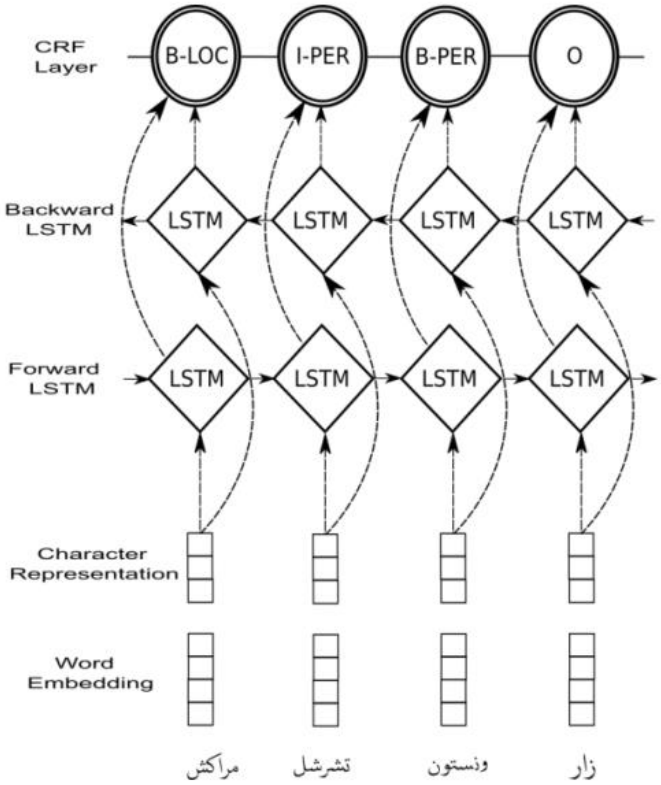
Pada jurnal keenam dengan judul "*Arabic Named Entity Recognition using deep learning approach"*, membahas NER berbahasa Arab yang sangat bergantung pada sumber daya eksternal dan rekayasa fitur untuk menghasilkan *state-of-the-art*. Untuk

mengatasi keterbatasan tersebut, diusulkan untuk menggunakan pendekatan *deep learning* untuk menangani tugas NER bahasa Arab. Diperkenalkan juga *neural network architecture* dengan dua arah *Long Short-Term Memory* (LSTM) dan *Conditional Random Fields* (CRF) dan diujicobakan dengan berbagai *hyperparameter* yang umum digunakan untuk menilai pengaruhnya terhadap kinerja keseluruhan sistem. Model yang dibangun mendapatkan dua sumber informasi tentang kata-kata sebagai masukan yaitu kata yang telah dilatih sebelumnya dan representasi berbasis karakter, serta menghilangkan kebutuhan akan pengetahuan khusus atau rekayasa fiturnya. Hasilnya diperoleh *state-of-the-art* pada corpus ANERcorp dengan standar skor F1 90,6%.

Pendekatan *deep learning* yang digunakan untuk menangani tugas NER untuk bahasa Arab digunakan arsitektur jaringan saraf yang terdiri dari lapisan BiLSTM dan lapisan CRF. Pertama, menghitung representasi karakter untuk setiap kata menggunakan CNN atau BiLSTM seperti gambar 11, kemudian menggabungkannya dengan kata yang di *embedding* sebelum dimasukkan ke lapisan BiLSTM. Lapisan ini terdiri dari dua jaringan LSTM. *Forward* LSTM membaca urutan kata dari awal dan ketika *backward* LSTM membacanya dalam urutan yang berlawanan. Vektor keluaran dari kedua jaringan LSTM digabungkan dan dikirim sebagai masukan ke lapisan CRF untuk menghasilkan prediksi tag untuk urutan masukan. Arsitektur jaringan saraf diilustrasikan secara rinci seperti pada gambar 12.



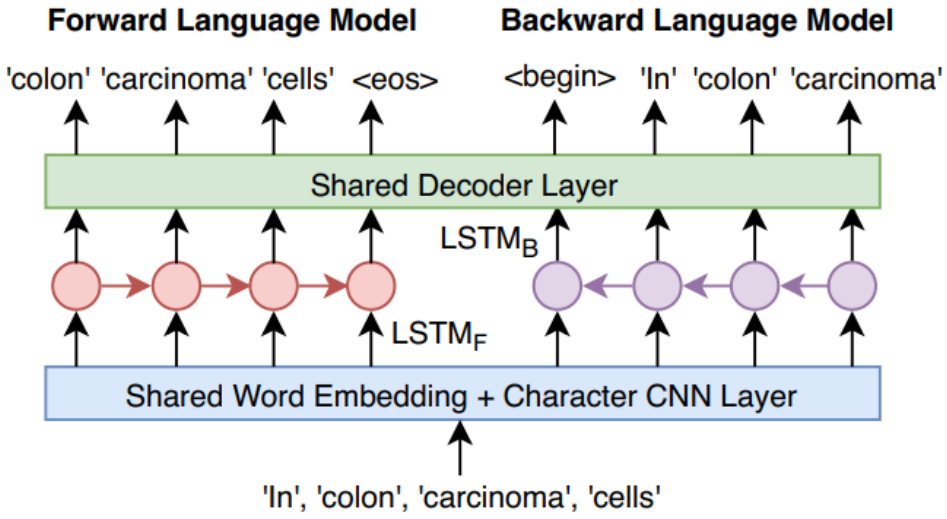
Gambar 11. Representasi karakter dasar menggunakan CNN



Gambar 12. Arsitektur utama jaringan saraf

Pada pembahasan jurnal ketujuh dengan judul "*Effective Use of Bidirectional Language Modeling for Transfer* Learning *in Biomedical Named Entity Recognition",*

NER biomedis menjadi konsep dasar dalam penambangan teks dokumen medis dan memiliki banyak aplikasi. Pendekatan berbasis *deep leaning* untuk tugas ini telah mendapatkan perhatian yang meningkat dalam beberapa tahun terakhir karena parameternya dapat dipelajari *end to end* tanpa campur tangan manusia. Namun, pendekatan ini mengandalkan *high-quality labeled data* dan hal ini sangat mahal untuk diperoleh. Untuk mengatasi masalah tersebut dilakukan penelitian untuk mendapatkan cara menggunakan *unlabeled text data* untuk meningkatkan kinerja model NER.

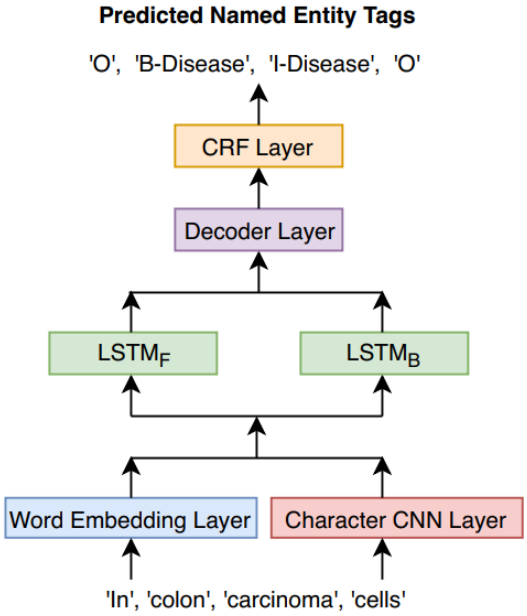


Gambar 13. Arsitektur BiLM

Secara khusus digunakan *bidirectional language model* (BiLM) seperti gambar 13 untuk melatih data tak berlabel (*unlabeled text data*) dan mentransfer bobotnya untuk melatih model NER dengan arsitektur yang sama seperti BiLM dan menghasilkan inisialisasi parameter yang lebih baik dari model NER. Selain itu juga dilakukan evaluasi pendekatan pada empat kumpulan data tolok ukur untuk NER biomedis dan menunjukkan bahwa hal itu mengarah pada peningkatan substansial dalam skor F1 dibandingkan dengan pendekatan *state-of-the-art* nya. Disini dapat ditunjukkan bahwa penggunaan BiLM akan mengarah ke pelatihan model yang lebih cepat dan model yang dilatih sebelumnya memerlukan lebih sedikit contoh data latih untuk mencapai skor F1 tertentu.

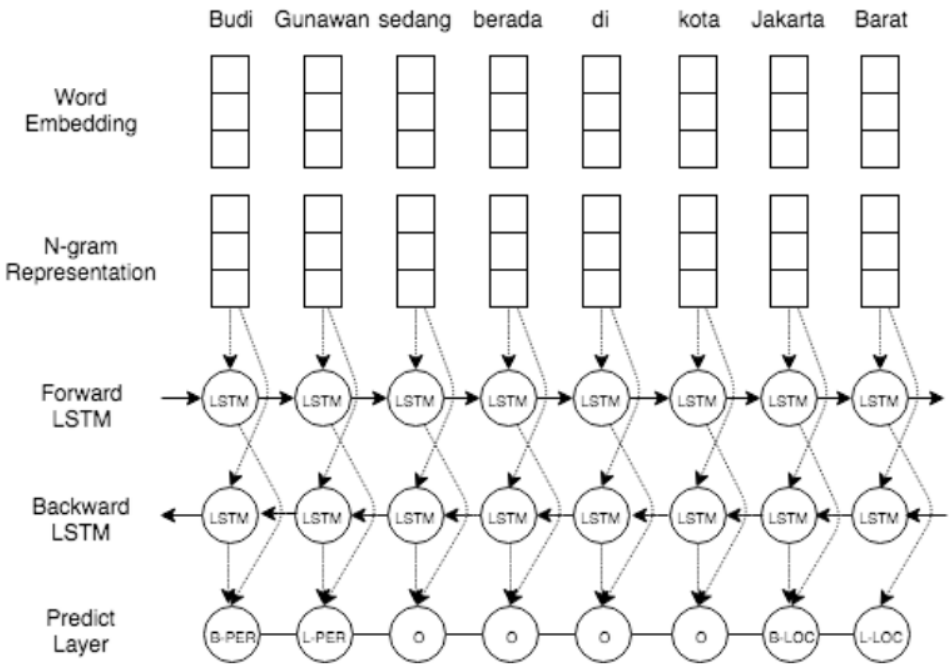
Design utama dari model NER berbasis jaringan saraf adalah lapisan jaringan *convolutional neural network* (CNN), lapisan *word embedding*, level kata layer BiLSTM, lapisan dekoder dan lapisan prediksi label pada tingkat kalimat (gambar 14). Selama pelatihan model, semua lapisan dilatih bersama. Sebelum pelatihan, juga

dilatih terlebih dahulu parameter lapisan karakter-CNN, *word embedding* dan BiLSTM dalam model NER menggunakan parameter yang dipelajari dari model bahasa yang memiliki arsitektur yang sama.



Gambar 14. Arsitektur Model NER

Pembahasan pada jurnal kedelapan dengan judul "*Named-Entity Recognition for Indonesian Language using Bidirectional LSTM-CNNs"* dijelaskan implementasi NER untuk Bahasa Indonesia dengan menggunakan berbagai pendekatan *deep learning*, namun terutama difokuskan pada arsitektur hybrid Bidirectional LSTM (BLSTM) dan *Convolutional Neural Network* (CNN). Sudah ada beberapa NER untuk pengembangan beberapa bahasa, antara lain bahasa Inggris, Vietnam, Jerman, India dan banyak lagi lainnya. Namun penelitian pada jurnal ini berfokus pada bahasa Indonesia. NER bahasa Indonesia ini berhasil mengekstrak informasi dari artikel ke dalam 4 (empat) kelas yang berbeda, yaitu Orang, Organisasi, Lokasi, dan Kejadian. Diberikan perbandingan komprehensif di antara semua eksperimen dengan menggunakan pendekatan *deep learning*.



Gambar 15. Arsitektur Jaringan Saraf menggunakan gabungan BSLTM dan CNN

Seperti yang dijelaskan pada gambar 15, arsitektur dimulai dari lapisan *word embedding*. Tahap kedua, representasi N-gram dibangun dan diolah dengan menggunakan layer CNN. Penggunaan metode BLSTM berarti bahwa model LSTM dibangun dalam dua arah, *forward* dan *backward*. Lapisan ini dipasang setelah lapisan CNN. Akhirnya, lapisan yang sepenuhnya terhubung diletakkan untuk memprediksi hasilnya.

Pada jurnal kesembilan dengan judul "*An Analysis of the Performance of Named Entity Recognition over OCRed Documents"* dibahas terkait penggunaan perpustakaan digital yang membutuhkan kemudahan akses ke dokumen yang sangat dipengaruhi oleh kualitas pengindeksan dokumen. *Named* Entity adalah salah satu informasi terpenting untuk mengindeks dokumen digital. Menurut studi terbaru, 80% dari 500 *query* teratas yang dikirim ke portal perpustakaan digital berisi setidaknya satu *Named Entity*. Namun sebagian besar dokumen digital diindeks melalui versi OCRed yang terjadi banyak kesalahan sehingga dapat menghalangi akses ke dokumen tersebut.

Named Entity Recognition (NER) adalah proses yang bertujuan untuk menemukan nama-nama penting dalam teks tertentu dan mengkategorikannya ke dalam satu set kelas yang telah ditentukan (orang, lokasi, organisasi). Jurnal tersebut bertujuan untuk memperkirakan kinerja sistem NER melalui data OCRed. Secara mendalam dibahas kesalahan NER yang timbul dari kesalahan OCR dan mempelajari

korelasi antara akurasi NER dengan tingkat kesalahan OCR serta memperkirakan biaya penyisipan karakter, penghapusan dan substitusi dalam *Named Entity*. Hasilnya menunjukkan bahwa meskipun mesin OCR mempengaruhi *Named Entity* dengan kesalahan, sistem NER dapat mengatasi masalah ini dan mengenali beberapa di antaranya dengan benar.

Table 2. NER Performance dan Hasil OCR

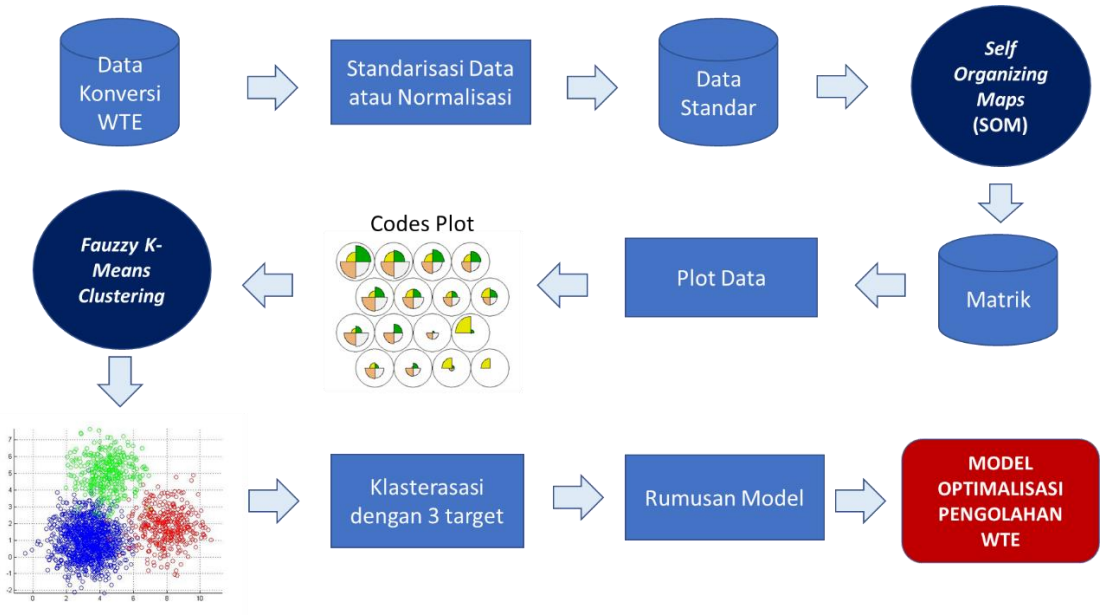
	OCR			NER		
	CER	WER	ENER	Pre	Rec	F1-score
Clean	––	––	––	89.4	90.8	90.1
LEV-0	1.7	8.5	6.9	83.7	90.7	86.8
Bleed	1.8	8.6	7.1	84.0	84.1	84.1
PhantChar	1.7	8.8	7.8	75.8	78.6	77.1
Blurring	6.3	20.0	21.5	66.9	69.5	68.8
CharDeg	3.6	21.8	23.4	64.5	64.8	64.7

Table 2 menunjukkan hasil NER pada *clean text* dan OCR dengan tingkat kesalahan OCR, selain *Character Error Rate* (CER) dan *Word Error Rate* (WER) juga ditampilkan hasil proporsi *Named Entity* yang salah yang diekstraksi oleh sistem OCR (ENER).

BAB III
METODE PENELITIAN

3.1 Skema Penelitian

Skema penelitian bawah ini yang akan dijadikan dasar sebagai alur penelitian agar dapat dicapai sebuah tujuan penelitian sesuai dengan target dan tidak menyimpang dari judul penelitian. Adapun skema penelitian seperti gambar dibawah ini.



Gambar 16. Skema Penelitian

Pada penelitian ini akan dilakukan beberapa tahapan sebagai berikut :

1. Data Konversi WTE adalah hasil pengujian dan pengukuran proses WTE dengan jumlah minimal 150 kombinasi dilakukan standarisasi atau normalisasi data.
2. Hasil data yang telah terstandar diproses dengan metode Self Organizing Maps (SOM) dengan hasil data matrik.
3. Selanjutnya diproses plot data untuk menghasilkan image Codes Plot yang selanjutnya diolah kembali dengan FCM atau Hirarchical Clustering Method untuk menghasilkan clasterisasi seuai kelompok target capaian kalori.
4. Hasil pengklasifikasian akan dirumuskan menjadi sebuah model optimalisasi pengolahan WTE yang dapat disesuaikan dengan karakteristik sampah yang berbeda-beda di seluruh Indonesia.

3.2 Pengambilan Data Konversi WTE

Data konversi pada proses WTE merupakan sebuah data kombinasi yang terdiri dari beberapa variabel yaitu durasi (hari), prosentase komposisi sampah organik dan non organik, konsentrat bioaktifator dan kalori hasil dari komposisi tersebut.

Pengambilan data dilakukan selama 6 (enam) bulan terkait dengan jumlah kombinasi dan lamanya hari yang ditetapkan terhadap setiap kombinasi. Hasil akhir dari proses “peuyemisasi” untuk menghasilkan matrial dalam bentuk pelet akan diuji menggunakan boom kalorimeter untuk mendapatkan hasil nilai kalori dari setiap matrial yang dikasilkan.

3.3 Standarisasi Data Konversi

Standarisasi data konversi dilakukan setelah didapatkan data-data kombinasi sehingga diharapakan akan didapat data yang lebih sederhana.

3.4 Self Organizing Maps (SOM)

Langkah selanjutnya dilakukan SOM terhadap data-data konversi yang diberi nama *som.wte* dimana data konversi akan diubah menjadi matrik dengan perintah :

```
som.wte <- som(as.matrix(wte[,-5]), grid = somgrid(xdim = 4, ydim = 4,
topo="hexagonal"))
```

3.5 Ploting Data Matriks

Data matrik yang dihasilkan dari proses SOM dilakukan proses ploting data sehingga akan dihasilkan codes plot untuk dapat memastikan bahwa keempat variabel tidak ada yang tereduksi atau dihilangkan.

3.6 Klasterisasi Fuzzy C-Means

Proses klasterisasi menggunakan FCM dan SOM digunakan untuk mendapatkan dan memastikan satu set profil pengolahan sampah mejadi bahan baku energi (breket) dengan kelompok kombinasi yang sesuai dengan hasil potensi kalori yang merujuk pada kombinasi variabel tersebut.

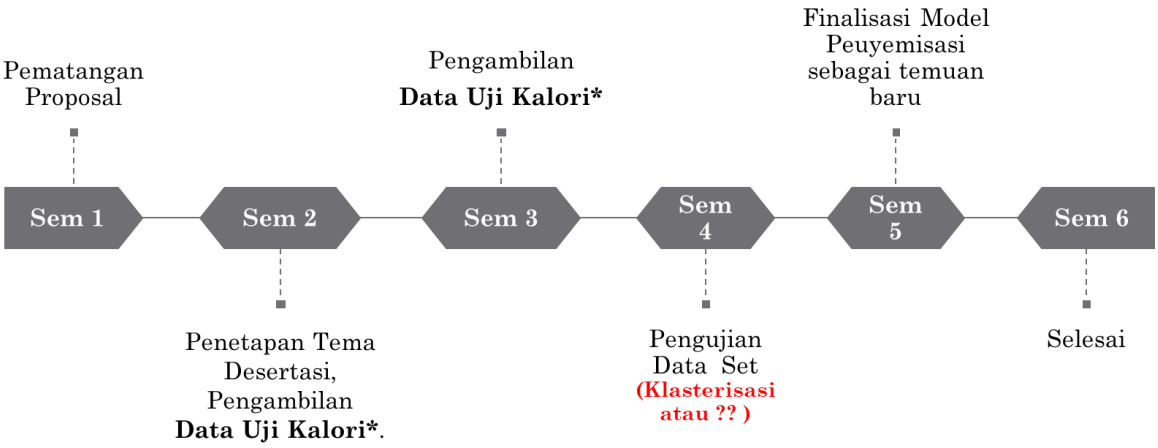
3.7 Perumusan Model

Perumusan model menjadi langkah terakhir untuk menetapkan kombinasi yang optimum dalam pengelolaan sampah menjadi energi (WTE). Model ini sangat penting

karena akan menjadi sebuah **standar baru** dalam pengelolaan sampah menjadi energi disetiap lokasi/ wilayah yang memiliki kombinasi sampah yang berbeda-beda. Dengan pengolahan data-data yang ada dalam rangkaian proses WTE akan diperoleh sebuah pengeloaan WTE yang optimum disetiap wilayah sehingga selanjutnya dapat diperoleh prediksi target waktu tercepat capaian bauran EBT di Indonesia khususnya dari proses WTE.

3.8 Rencana Kegiatan

Untuk mencapai target penelitian/ desertasi, maka penulis menyusun rencana kegiatan berupa jadwal kegiatan yang berguna untuk memastikan agar capaian yang ditetapkan dapat dipenuhi sesuai waktu yang telah ditetapkan. Adapun jadwal yang akan digunakan sebagai beikut :



Gambar 17. Jadwal Desertasi

REFERENCES

Albores, P., Petridis, K., & Dey, P. K. (2016). Analysing Efficiency of Waste to Energy Systems: Using Data Envelopment Analysis in Municipal Solid Waste Management. *Procedia Environmental Sciences*, *35*, 265–278. https://doi.org/10.1016/j.proenv.2016.07.007

El Bazi, I., & Laachfoubi, N. (2019). Arabic named entity recognition using deep learning approach. *International Journal of Electrical and Computer Engineering*, *9*(3), 2025–2032. https://doi.org/10.11591/ijece.v9i3.pp2025-2032

Filannino, M., Brown, G., & Nenadic, G. (2013). ManTIME: Temporal expression identification and normalization in the TempEval-3 challenge. **SEM 2013 - 2nd Joint Conference on Lexical and Computational Semantics*, *2*, 53–57.

Gunawan, W., Suhartono, D., Purnomo, F., & Ongko, A. (2018). Named-Entity Recognition for Indonesian Language using Bidirectional LSTM-CNNs. *Procedia Computer Science*, *135*, 425–432. https://doi.org/10.1016/j.procs.2018.08.193

Kim, J., Ko, Y., & Seo, J. (2019). A Bootstrapping Approach with CRF and Deep Learning Models for Improving the Biomedical Named Entity Recognition in Multi-Domains. *IEEE Access*, *7*, 70308–70318. https://doi.org/10.1109/ACCESS.2019.2914168

Li, M. J., Ng, M. K., Cheung, Y. M., & Huang, J. Z. (2008). Agglomerative fuzzy K-Means clustering algorithm with selection of number of clusters. *IEEE Transactions on Knowledge and Data Engineering*, *20*(11), 1519–1534. https://doi.org/10.1109/TKDE.2008.88

Nakagawa, K., Amano, H., Kawamura, A., & Berndtsson, R. (2017). Classification of groundwater chemistry in Shimabara, using self-organizing maps. *Hydrology Research*, *48*(3), 840–850. https://doi.org/10.2166/nh.2016.072

Niska, H., & Serkkola, A. (2018). Data analytics approach to create waste generation profiles for waste management and collection. *Waste Management*, *77*, 477–485. https://doi.org/10.1016/j.wasman.2018.04.033

Sachan, D. S., Xie, P., Sachan, M., & Xing, E. P. (2017). *Effective Use of Bidirectional Language Modeling for Transfer Learning in Biomedical Named Entity Recognition*. 1–19. http://arxiv.org/abs/1711.07908