TEXT MINING

TUGAS AKHIR PRAKTIKUM

Dzikri Faizziyan | 065118123 | Gabung C
Ananda Reynata Saputra | 065118180 | Gabung C

Ψ

- **PENJELASAN & TUJUAN TERKAIT TOPIK YANG DIPILIH**

  Dota 2 memiliki 121 karakter hero dengan kompleksitas dan role yang berbeda-beda sesuai dengan skill dan atribut yang mereka miliki. Sebagai Game Multiplayer yang memprioritaskan kerja sama antar pemain, Dota 2 mewajibkan tiap pemainnya untuk memilih pos dan role sesuai dengan keinginan dan kemampuan masing-masing pemain.Terdapat 5 role utama dalam Dota 2 yaitu : *Carry, Midlaner, Offlaner, Soft Support dan Hard Support*. Secara umum, suatu tim dalam Dota 2 diisi oleh 3 Core dan 2 Support.

  Dalam Dota 2, semua posisi itu penting tetapi para hero yang memiliki posisi sebagai core ini merupakan role yang paling penting untuk membawa kemenangan bagi sebuah tim terutama pada game late atau ketika game berada pada tahap akhir. Untuk posisi support ini biasanya diisi oleh hero dengan offensive dan roaming yang baik, skill yang bisa menyelamatkan atau membunuh hero musuh dengan baik dan juga mereka biasanya membeli item yang berkontribusi untuk team. Posisi support ini bertugas hampir disemua lane untuk membantu hero pada lane tersebut mendapatkan kill dan mendapatkan Exp dan Gold lebih cepat. Untuk memenangkan game, tiap role ini harus saling bekerjasama dan kompak, apabila salah satu dari role ini pincang maka akan sulit untuk memenangkan game tersebut.

  Oleh sebab itu, pada kesempatan kali ini tujuan kami bermaksud akan mencoba mengklasifikasikan Role dari beberapa Hero yang ada di Dota 2 agar mempermudah untuk mengetahui mana Role yang termasuk Core dan mana Role yang termasuk Support

- **MEMBUAT CORPUS DENGAN FILE YANG BEREKTENSI XLSX**

  Tentunya sebelum mencoba masuk untuk mengolah atau memproses data, terlebih dahulu kita menyiapkan data yang akan kita proses dan kita oleh nantinya, disini kami membuat 2 corpus dengan format file xlsx, Corpus tersebut terdiri dari :

  1. Data yang berisi 20 file dokumen (Sebagai Train Test), dan
  2. Data yang berisi 12 file dokumen (Sebagai Predict Test).

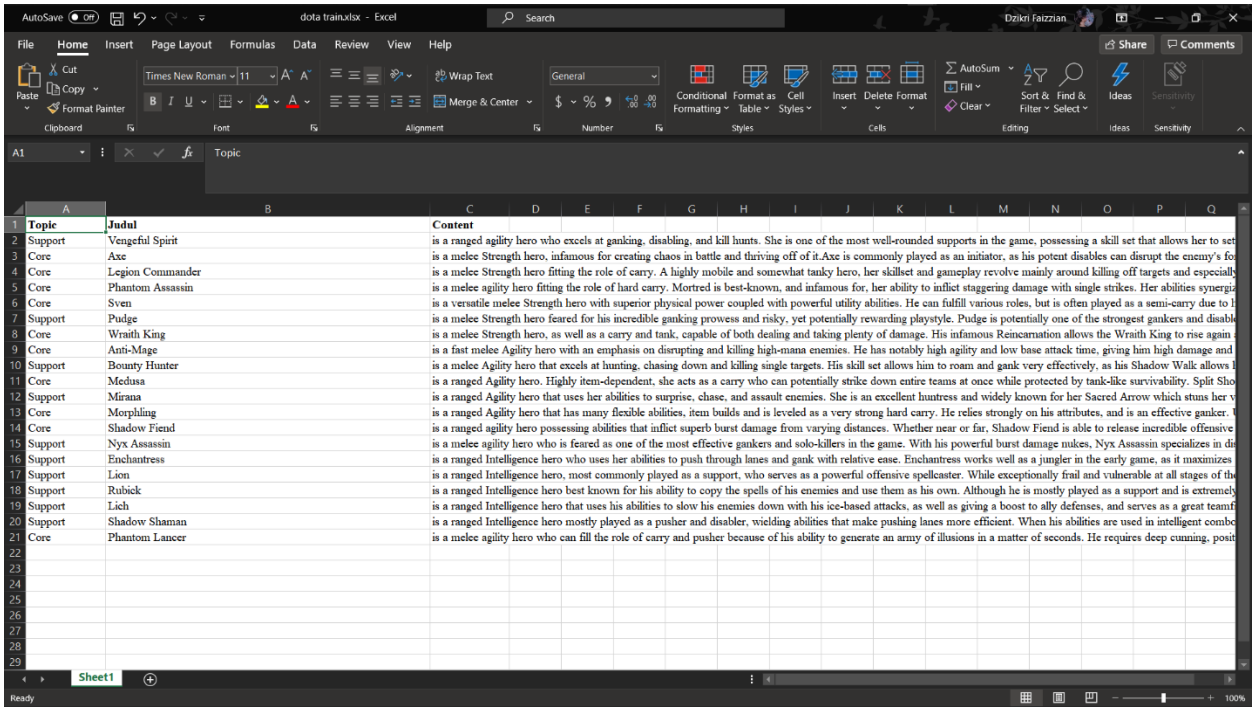
  Deskripsi/penjelasan terkait isi didalam corpus/data tersebut meliputi :

  1. Topic : Sebagai Klasifikasi role dari beberapa Hero Dota 2
  2. Judul : Sebagai Nama Karakter dari Role/Hero Dota 2
  3. Content : Sebagai Deskripsi dari Karakter Role/Hero Dota 2

  Tetapi pada bagian data Predict Test hanya mencakup Judul dan Content saja, yang tentunya ini sebagai tujuan kami untuk mengklasifikasikan Role dari beberapa Hero yang ada di Dota 2 tersebut.

Berikut adalah Corpus/data kami yang akan kami coba olah dan klasifikasikan dengan menggunakan software Orange for Data Mining.

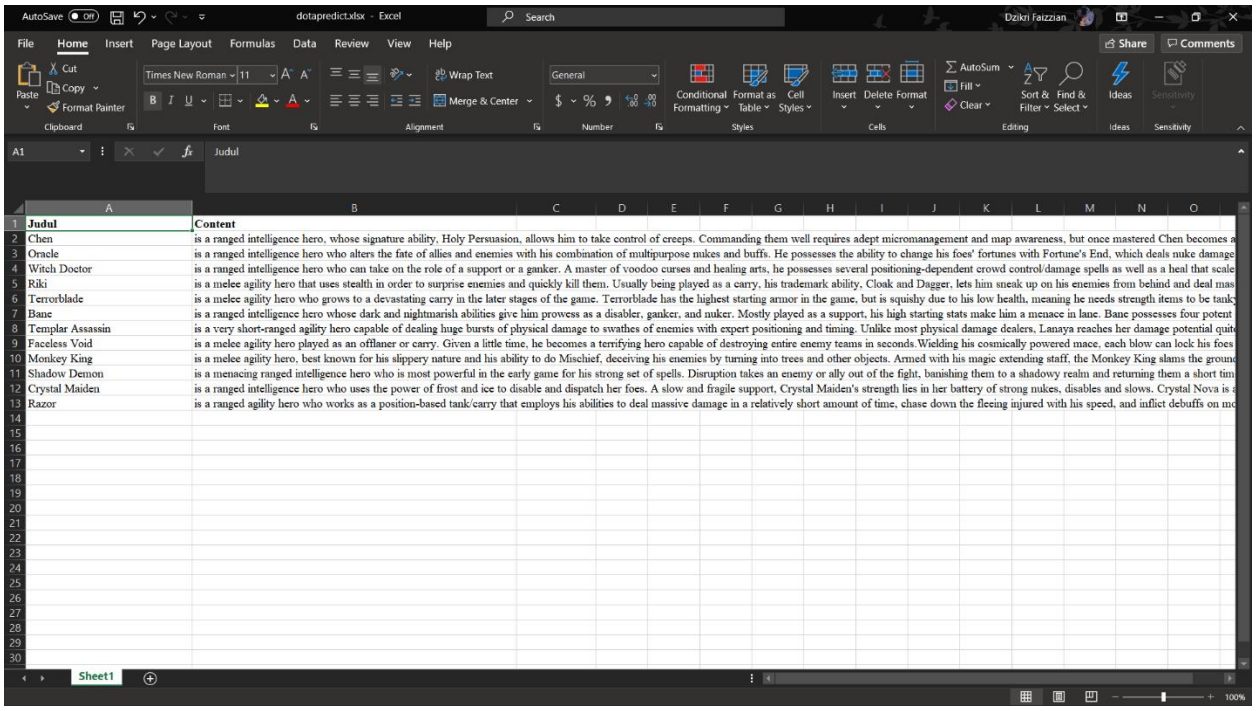
- Data yang berisi 20 file dokumen (Sebagai Train Test).



Topic	Judul	Content
Support	Vengeful Spirit	is a ranged agility hero who excels at ganking, disabling, and kill hunts. She is one of the most well-rounded supports in the game, possessing a skill set that allows her to set
Core	Axe	is a melee Strength hero, infamous for creating chaos in battle and thriving off of it.Axe is commonly played as an initiator, as his potent disables can disrupt the enemy's fo
Core	Legion Commander	is a melee Strength hero fitting the role of carry. A highly mobile and somewhat tanky hero, her skillset and gameplay revolve mainly around killing off targets and especiall
Core	Phantom Assassin	is a melee agility hero fitting the role of hard carry. Mortred is best-known, and infamous for, her ability to inflict staggering damage with single strikes. Her abilities synergiz
Core	Sven	is a versatile melee Strength hero with superior physical power coupled with powerful utility abilities. He can fulfill various roles, but is often played as a semi-carry due to h
Support	Pudge	is a melee Strength hero feared for his incredible ganking prowess and risky, yet potentially rewarding playstyle. Pudge is potentially one of the strongest gankers and disabl
Core	Wraith King	is a melee Strength hero, as well as a carry and tank, capable of both dealing and taking plenty of damage. His infamous Reincarnation allows the Wraith King to rise again a
Core	Anti-Mage	is a fast melee Agility hero with an emphasis on disrupting and killing high-mana enemies. He has notably high agility and low base attack time, giving him high damage and
Support	Bounty Hunter	is a melee Agility hero that excels at hunting, chasing down and killing single targets. His skill set allows him to roam and gank very effectively, as his Shadow Walk allows h
Core	Medusa	is a ranged Agility hero. Highly item-dependent, she acts as a carry who can potentially strike down entire teams at once while protected by tank-like survivability. Split Sho
Support	Mirana	is a ranged Agility hero that uses her abilities to surprise, chase, and assault enemies. She is an excellent huntress and widely known for her Sacred Arrow which stuns her v
Core	Morphling	is a ranged Agility hero that has many flexible abilities, item builds and is leveled as a very strong hard carry. He relies strongly on his attributes, and is an effective ganker.
Core	Shadow Fiend	is a ranged agility hero possessing abilities that inflict superb burst damage from varying distances. Whether near or far, Shadow Fiend is able to release incredible offensive
Support	Nyx Assassin	is a melee agility hero who is feared as one of the most effective gankers and solo-killers in the game. With his powerful burst damage nukes, Nyx Assassin specializes in di
Support	Enchantress	is a ranged Intelligence hero who uses her abilities to push through lanes and gank with relative ease. Enchantress works well as a jungler in the early game, as it maximizes
Support	Lion	is a ranged Intelligence hero, most commonly played as a support, who serves as a powerful offensive spellcaster. While exceptionally frail and vulnerable at all stages of th
Support	Rubick	is a ranged Intelligence hero best known for his ability to copy the spells of his enemies and use them as his own. Although he is mostly played as a support and is extremely
Support	Lich	is a ranged Intelligence hero that uses his abilities to slow his enemies down with his ice-based attacks, as well as giving a boost to ally defenses, and serves as a great teamf
Support	Shadow Shaman	is a ranged Intelligence hero mostly played as a pusher and disabler, wielding abilities that make pushing lanes more efficient. When his abilities are used in intelligent combo
Core	Phantom Lancer	is a melee agility hero who can fill the role of carry and pusher because of his ability to generate an army of illusions in a matter of seconds. He requires deep cunning, positi

Dota 2 Train Test

- Data yang berisi 12 file dokumen (Sebagai Predict Test).



Judul	Content
Chen	is a ranged intelligence hero, whose signature ability, Holy Persuasion, allows him to take control of creeps. Commanding them well requires adept micromanagement and map awareness, but once mastered Chen becomes a
Oracle	is a ranged intelligence hero who alters the fate of allies and enemies with his combination of multipurpose nukes and buffs. He possesses the ability to change his foes' fortunes with Fortune's End, which deals nuke damage
Witch Doctor	is a ranged intelligence hero who can take on the role of a support or a ganker. A master of voodoo curses and healing arts, he possesses several positioning-dependent crowd control/damage spells as well as a heal that scale
Riki	is a melee agility hero that uses stealth in order to surprise enemies and quickly kill them. Usually being played as a carry, his trademark ability, Cloak and Dagger, lets him sneak up on his enemies from behind and deal mas
Terrorblade	is a melee agility hero who grows to a devastating carry in the later stages of the game. Terrorblade has the highest starting armor in the game, but is squishy due to his low health, meaning he needs strength items to be tanky
Bane	is a ranged intelligence hero whose dark and nightmarish abilities give him prowess as a disabler, ganker, and nuker. Mostly played as a support, his high starting stats make him a menace in lane. Bane possesses four potent
Templar Assassin	is a very short-ranged agility hero capable of dealing huge bursts of physical damage to swathes of enemies with expert positioning and timing. Unlike most physical damage dealers, Lanaya reaches her damage potential quite
Faceless Void	is a melee agility hero played as an offlaner or carry. Given a little time, he becomes a terrifying hero capable of destroying entire enemy teams in seconds.Wielding his cosmically powered mace, each blow can lock his foes
Monkey King	is a melee agility hero, best known for his slippery nature and his ability to do Mischief, deceiving his enemies by turning into trees and other objects. Armed with his magic extending staff, the Monkey King slams the ground
Shadow Demon	is a menacing ranged intelligence hero who is most powerful in the early game for his strong set of spells. Disruption takes an enemy or ally out of the fight, banishing them to a shadowy realm and returning them a short tim
Crystal Maiden	is a ranged intelligence hero who uses the power of frost and ice to disable and dispatch her foes. A slow and fragile support, Crystal Maiden's strength lies in her battery of strong nukes, disables and slows. Crystal Nova is a
Razor	is a ranged agility hero who works as a position-based tank/carry that employs his abilities to deal massive damage in a relatively short amount of time, chase down the fleeing injured with his speed, and inflict debuffs on mo

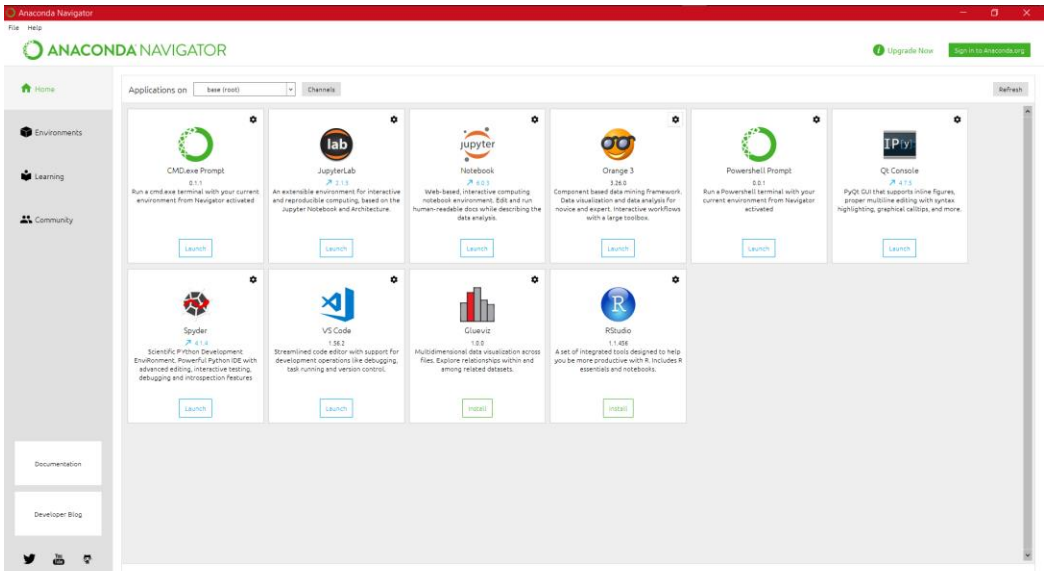
Dota 2 Predict Test

- **MENERAPKAN TEXT PROCESSING PADA DATA YANG AKAN DI OLAH.**

Sebelum kita masuk kedalam software orange untuk melakukan text preprocessing kita tentunya harus berkenalan terlebih dahulu terkait itu agar Ketika saat pengerjaan bisa mengerti terhadap apa yang sedang dilakukan. Tahap *pre-processing* atau praproses data merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Tentunya tujuan kami melakukan Praproses data itu untuk mempermudah dalam menyajikan data tersebut karena ini akan mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem.

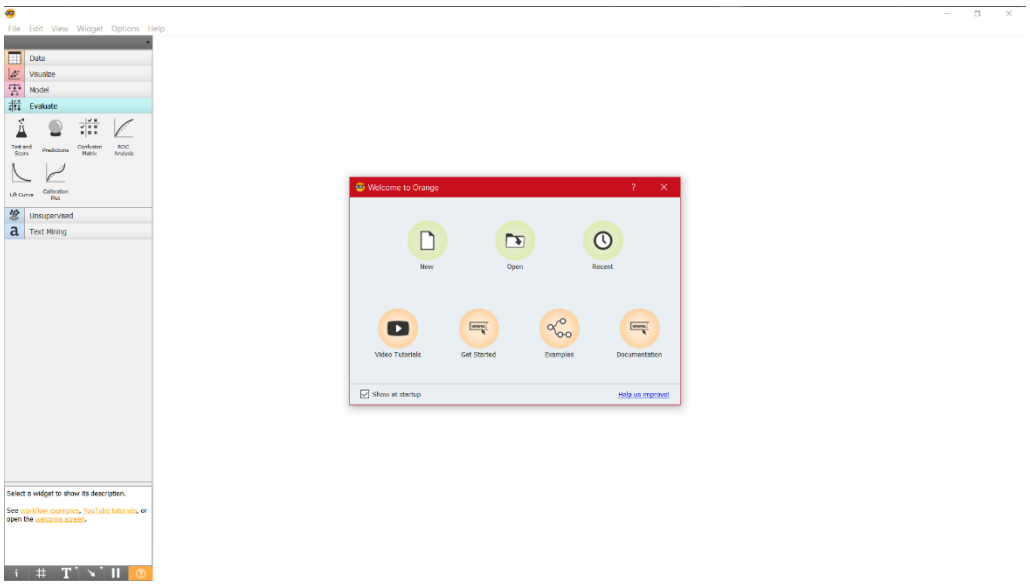
Baik untuk melakukan tahap preprocessing sebagai penjelasan tingkat lanjut, kita langsung saja membuka Aplikasi Orange for Data Mining terlebih dahulu,

Disini Kami menggunakan Ananconda Navigator untuk membuka Aplikasi Orange For Data Miningnya.



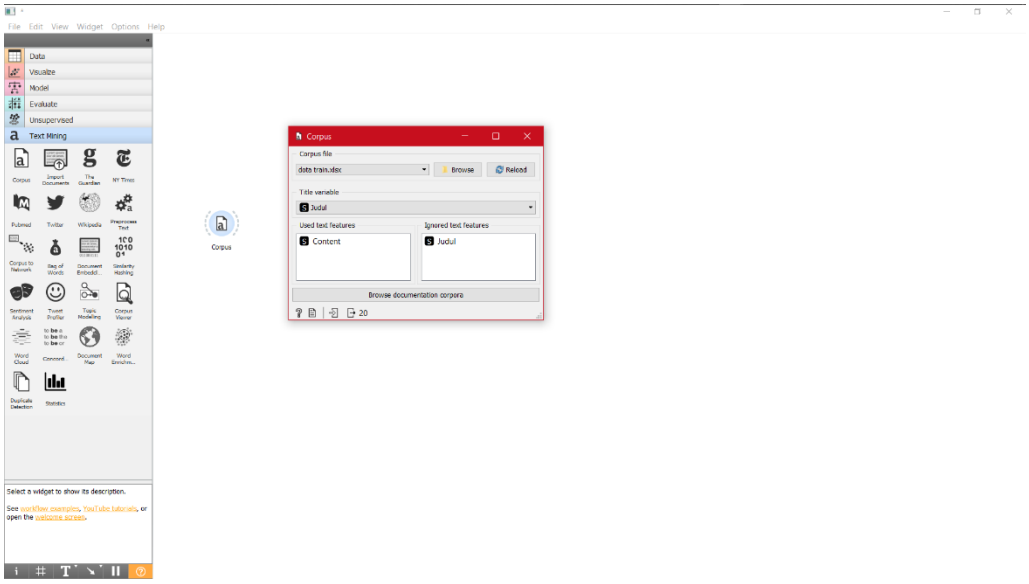
Anaconda Navigator

Lalu kita coba klik “Launch” pada tampilan Aplikasi Orange 3 for Data Miningnya, setelah itu kita klik “New” pada tampilan awal orange for data mining tersebut untuk mencoba melakukan pemrosesan data yang kita inginkan.

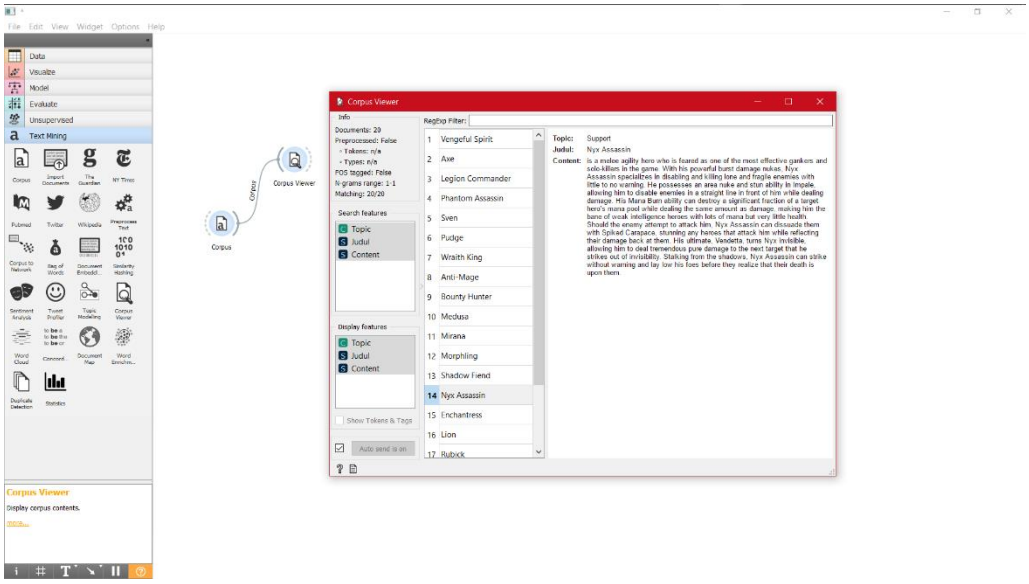


Tampilan Awal Aplikasi Orange For Data Mining

Selanjutnya kita klik widget yang bernama “Corpus” pada bagian Text Mining sebelah samping kiri atau bisa juga klik dan drag ke canvas. Selanjutnya double click Corpus. Maka tampilannya akan seperti dibawah ini :

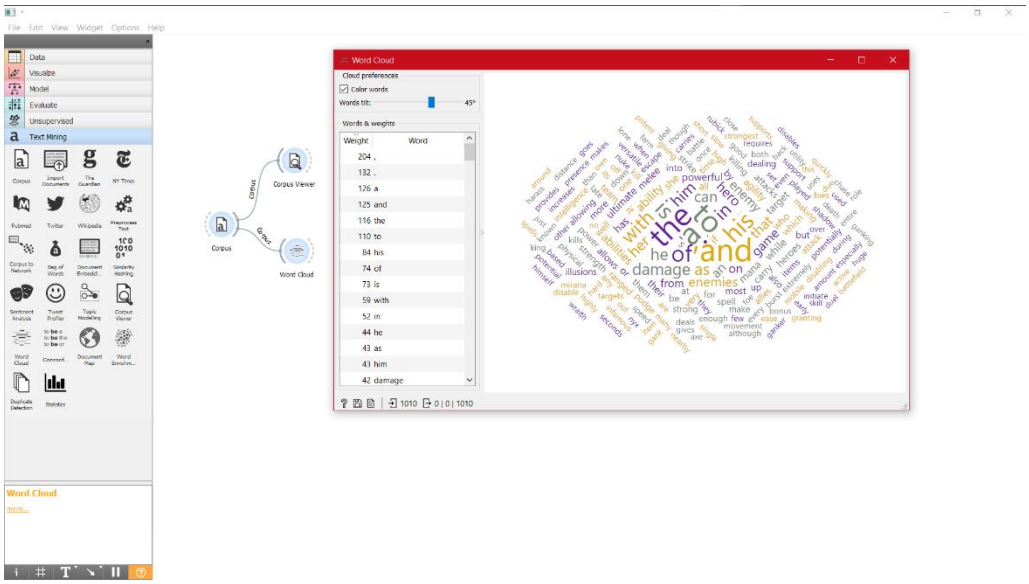


Pada bagian Corpus File terdapat 5 corpus yang disertakan secara default oleh aplikasi Orange, tetapi disini kami memilih “Browse” untuk memasukan data/corpus file yang sudah kami buat. Setelah berhasil memilih file silahkan pilih widget yang bernama “Corpus Viewer” untuk menampilkan isi dari corpus yang telah dipilih. Tarik dan hubungkan Corpus ke Corpus Viewer. Maka tampilannya akan menjadi seperti dibawah ini :



Pada Corpus Viewer ini kita diberikan informasi bahwa dalam File Corpus **Dota Train** yang sebelumnya kita pilih, ternyata terdapat 20 dokumen. Pada Corpus Viewer ada yang Namanya *Preprocessed: False* yang artinya kita belum melakukan tahap Preproses Teks. Pada *Display Features* → *Content* disini kalian bisa klik untuk menampilkan konten/dokumen yang ada pada corpus **Dota Train**. Misalnya Ketika kita memilih dokumen ***Nyx Assasin*** kita bisa melihat penjelasan/deskripsi dari kontennya yang ada di sebelah kanan. Seperti yang terlihat gambar diatas.

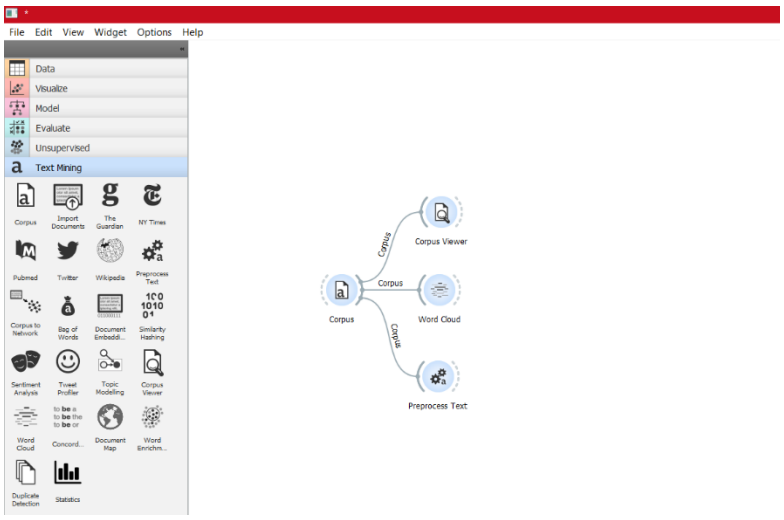
Setelah itu untuk melihat makna kata ataupun kata yang sering muncul dari file dokumen/corpus dota train bisa menggunakan salah satu widget yang bernama “Word Cloud” yang masih terletak pada bagian Text Mining, silahkan anda klik dan drag, lalu hubungkan dari Corpus ke Word Cloud setelah itu klik 2x pada Word Cloud, maka tampilannya akan seperti dibawah ini :



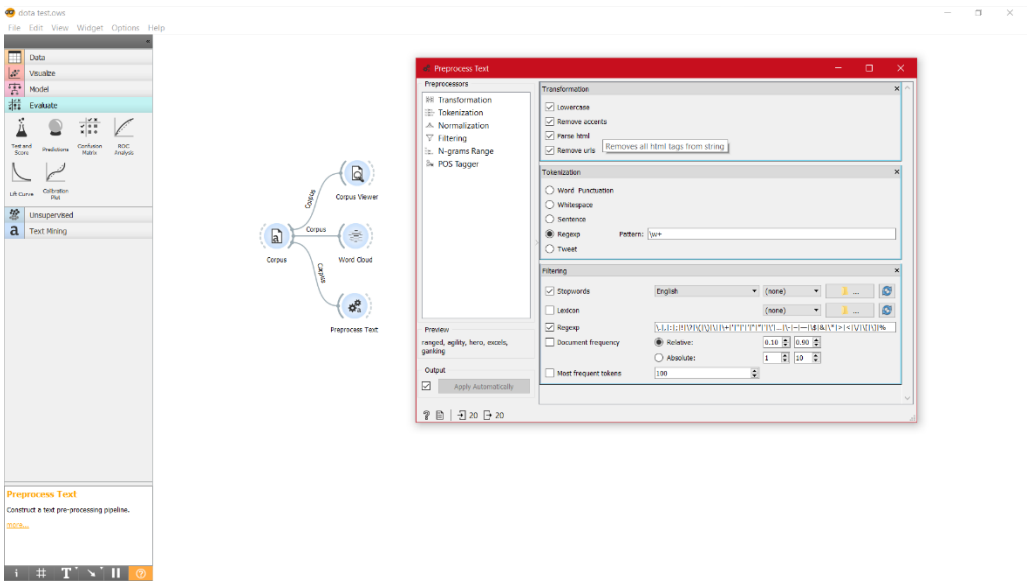
Disini kita dapat melihat bahwa kata yang paling bermakna dari Corpus **Dota Train** adalah tanda koma (,), Titik (.), a, and, the, to, his, dll.
Catatan : Font yang paling besar yang ditampillkan disamping adalah kata yang mempunyai bobot tertinggi.
Penjelasan pada Word Cloud :

- **Cloud Preferences** : Pada Color Word ini kita bisa mengatur apakah akan diberikan Warna ataupun tidak untuk word cloud kita dengan cara menceklis kotak yang sediakan disana.
- **Words Tilt** : Berfungsi mengatur ukuran dari word cloud, apakah mau berukuran besar ataupun ukuran kecil dengan menggeser slidebar yang telah disediakan.
- **Words & Weights** : Berfungsi untuk menampilkan bobot kata. Misalnya tanda koma (,) yang memiliki bobot sebesar 204 di korpus **Dota Train.**

Disini kita menyadari bahwa agak cukup mengganggu bukan untuk tanda koma, tanda titik, dan juga semacam kata penghubung semacam a, and, the, to, dll. Ini jelas cukup mengganggu dari performa algoritma dan menganggu sebagai pandangan secara estetika karena kurang elok untuk dilihat. Maka dari itu disini kita butuh yang disebut sebagai “NLP (Natural Language Processing)” terlebih dahulu singkatnya disebut sebagai “Text Processing”. Untuk melakukan itu kita coba pilih atau drag and drop Widget “Preprocessing Text” yang masih berada di bagian “Text Mining”. Lalu coba hubungkan Corpus ke Preprocessing Text. Seperti Tampilan dibawah ini :



Selanjutnya, coba kita double klik pada widget “Preprocess Text” untuk menampilkan pengaturan pada Preprocess Text. Maka tampilannya seperti dibawah ini :



Penjelasan singkat terkait fitur yang ada dalam Preprocess Text :

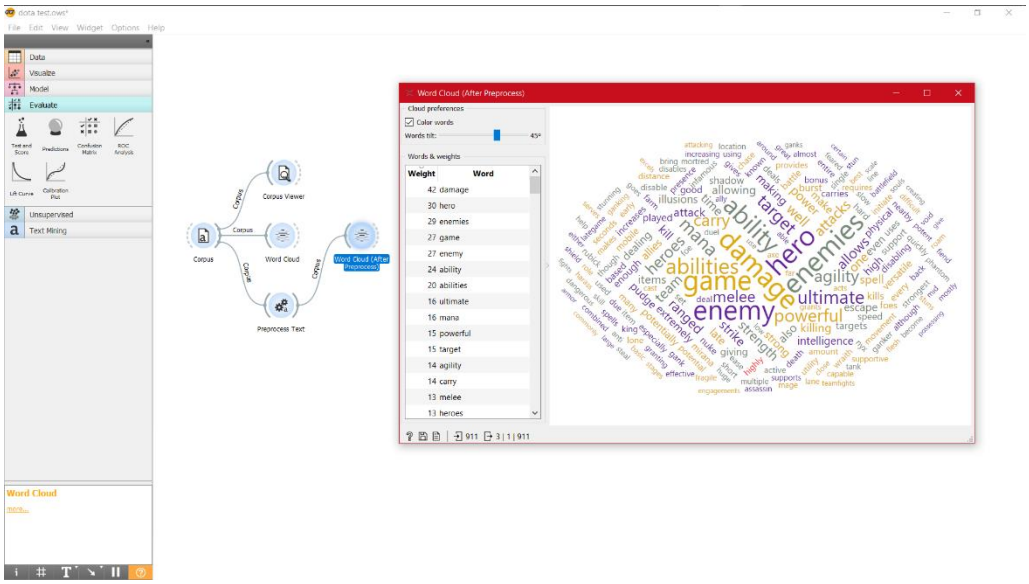
1. Pada Bagian Tranformation ini fungsinya untuk Mengubah teks. Dan disini juga kita bisa melakukan :
   - Lowercase : Mengubah teks ke huruf kecil.
   - Remove Accents: Menghilangkan Aksen dalam teks jika ada.
   - Parse HTML : Berfungsi jika dokumen kita ada HTML-nya kita bisa melakukan parsing.
   - Remove URL : menghapus link atau url jika ada dalam sebuah dokumen.
2. Pada Tokenization ini berfungsi untuk Melakukan tokenisasi dan beberapa preproses lain misalnya menghilangkan spasi/whitespace dan beberapa bagian dari teks dengan Regex.
3. Bagian Filtering tujuannya Memfilter dokumen dengan melakukan beberapa hal seperti berikut ini :
   - Stopwords : menghapus penghubung dalam teks. Kita bisa memilih bahasa sesuai dengan teks. Dan kita bisa menyisipkan stopwords sendiri dalam bentuk txt.
   - Lexicon : menghapus kosa kata
   - Regexp : Hapus tanda baca/ ekspresi Reguler.
   - Document Frequency : Menghapus frekuensi dokumen.
   - Most Frequent Tokens : Menghapus token yang ada.

Terlihat pada gambar diatas, pada bagian Tranformation kami mencentang semuanya, dari Lowercase, Remove Accent, Parse Html, dan Remove Url. Terutama khususnya kenapa kami mencentang pada bagian Remove Accent, dikarenakan dokumen yang kami gunakan adalah berbahasa Inggris, tentunya menggunakan fitur Remove Accent ini agar bisa menghilangkan Aksen dalam teks untuk mempermudah pembacaan dan tampilan pada dokumen tersebut.

Pada bagian Tokenization kami mencentang Regexp yang tujuannya untuk menghapus tanda baca atau ekspresi reguler yang ada pada dokumen tersebut.

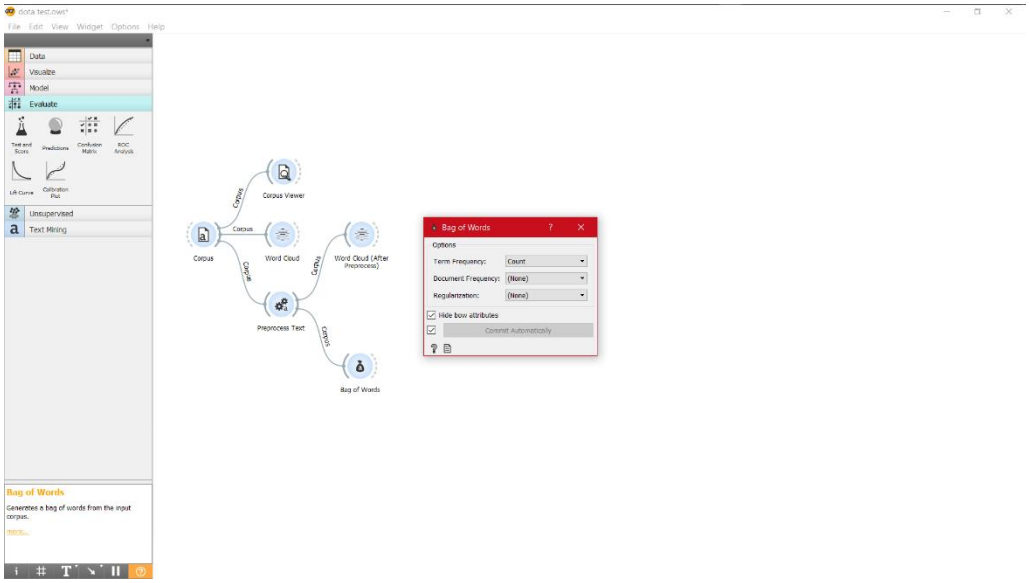
Dan pada bagian Filtering disini kami mencentang pada Stopword dan memilih English karena dokumen yang kami olah berbahasa inggris, Stopword ini berfungsi untuk menghapus penghubung dalam teks.

Setelah Tahap Preproccess Text selesai, kita coba tambahkan widget Word Cloud lagi. Lalu hubungkan Preprocessing Text ke Word Cloud, setelah itu double klik untuk menampilkannya. Tampilan akan seperti dibawah ini :



Setelah tahap prepocess text, tampilan pada dokumen tersebut lebih rapih dan enak dilihat bukan ? dan juga akan terlihat lebih jelas dari corpus Dota Train ini, seperti apa maknanya dari kata yang ada secara kasarnya. Terlihat juga disini kata “damage” adalah kata dengan bobot tertinggi kemudian ada kata “hero” dan “enemies”. Dan seterusnya.

Setelah itu coba tambahkan widget “Bag of Words.” Lalu hubungkan widget Preprocessing Text ke Bag of Words. Kita akan mengubah corpus yang telah dilakukan Preproses ke Bag of Words yang merepresentasikan setiap deskripsi dengan vektor dari word counts. Setelah coba dihubungkan silahkan Double click untuk melihat proses pada “Bag of Words”, didalamnya kita bisa melakukan opsi. Tampilannya akan seperti dibawah ini :



Penjelasan singkat terkait fitur yang ada dalam Bag of Word :

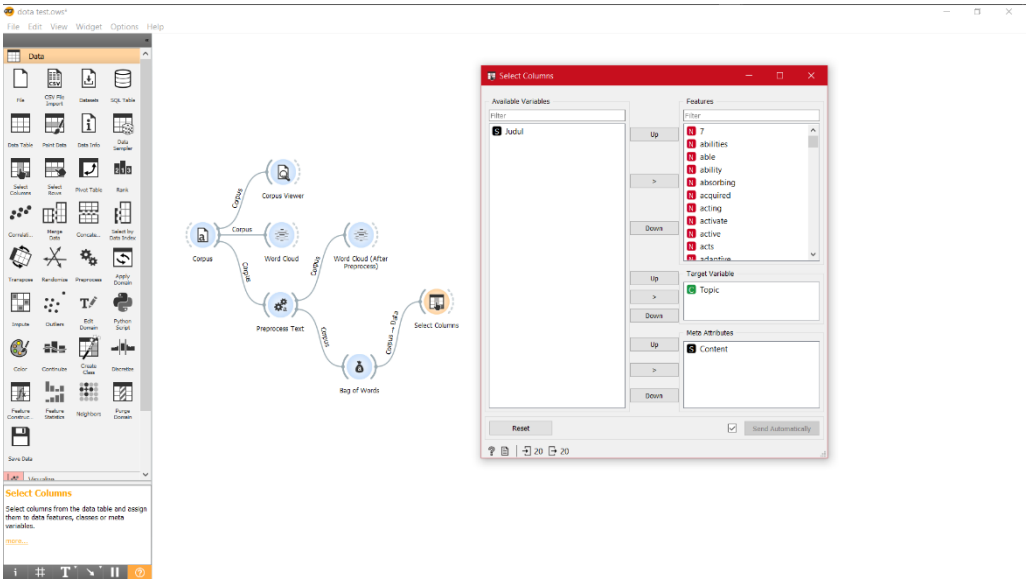
1. Term Frequency
   - Count : jumlah akurasi dari kata dalam dokumen
   - Binary : apakah kata dalam teks akan muncul atau tidak dalam dokumen
   - Sublinear : logaritma dari term frequency (perhitungan)
2. Document Frequency
   - None : tidak ada document frequency.
   - IDF : Inverse Document Frequency
   - Smooth-IDF : menambahkan satu ke document frequency untuk mencegah pembagian nol.
3. Regularization
   - L1 (Sum of Elements) : menormalisasi panjang vektor untuk penjumlahan dari elemen.
   - L2 (Euclidean) : menormalisasi panjang vektor ke jumlah kuadrat.

Pada bagian Bag of Word diatas, kami hanya memilih "Count" yang ada pada bagian "Term Frequency" yang fungsinya untuk menentukan jumlah akurasi dari kata dalam dokumen **Dota Train.**

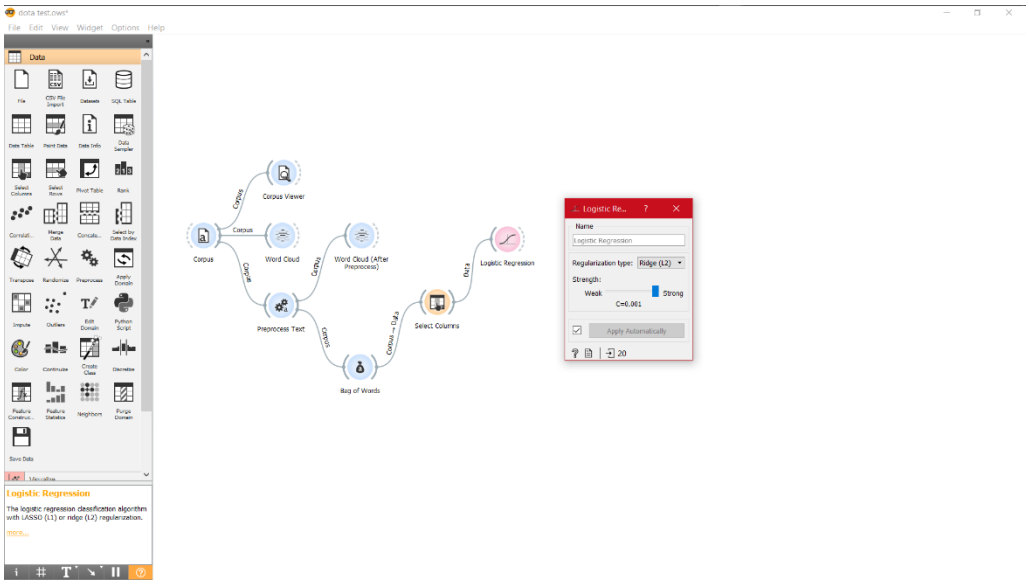
- **MENERAPKAN LOGISTIC REGRESSION PADA DATA YANG AKAN DI OLAH.**

Okay, tahap selanjutnya kita akan mencoba untuk mengimplementasikan Logistic Regression kedalam corpus Dota Train. Sebelumnya Kita harus berkenalan dulu dengan Logistic Regression.
Ap aitu Logistic Regression ?
Logistic Regression atau disebut Regresi Logistik mirip dengan Regresi Linier yang keduanya mempunyai variable dependent (variable pengaruh) dan independent (variable yang terpengaruh) yang biasanya dianalogikan dalam bentuk X dan Y. Keduanya sama-sama memiliki garis regresi, namun regresi logistik ini digunakan untuk memprediksi apakah sesuatu itu bernilai benar atau salah, ketimbang memprediksi suatu nilai yang kontinu seperti kasus yang menggunakan Regresi Linier (Arifin, 2018).
Jadi regresi logistik adalah teknik yang digunakan untuk memisahkan data menjadi dua bagian yaitu YA atau TIDAK / 0 atau 1. Teknik Logistic Regression biasanya digunakan untuk menyelesaikan masalah klasifikasi.

Nah Sebelum masuk kedalam Logistic Regression kita coba dahulu untuk menambah widget "Select Columns" pada widget add on Data, yang tujuannya agar lebih mudah mengkategorikan dan melihat pembagian data dari corpus untuk diproses pada Logistic Regression. Double Click maka tampilannya akan seperti dibawah ini :



Setelah itu untuk coba menerapkan Logistic Regression pada corpus, Caranya dengan mengarahkan ke widget/add on Model lalu pilih Logistic Regression. Dan coba hubungkan dari “Select Columns” ke “Logistic Regression”. Double Click pada Logistic Regression maka tampilannya akan seperti dibawah ini :

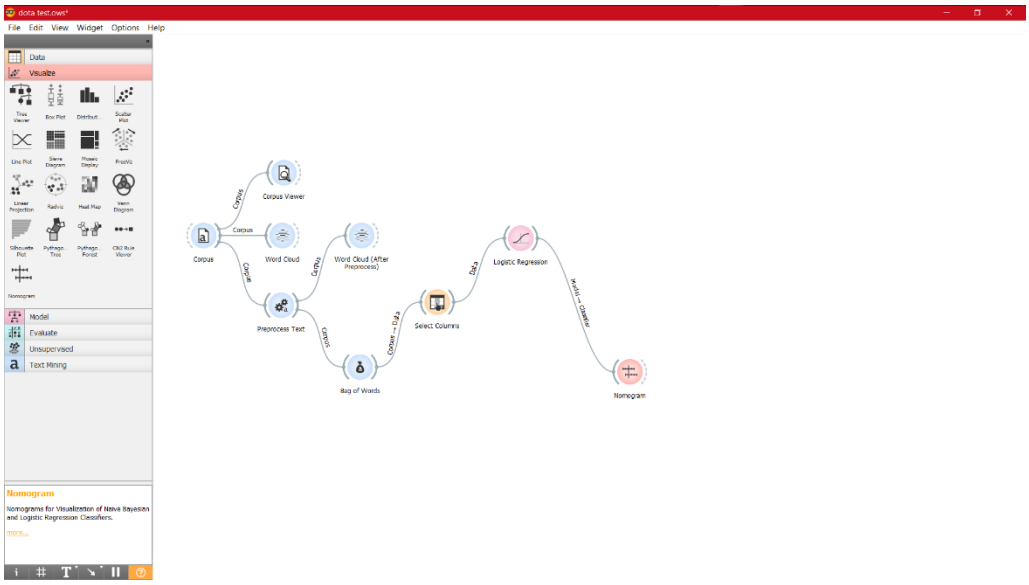


Disini pada Logistic Regression akan muncul beberapa opsi mulai dari mengganti nama, tipe regularisasinya, dan strength dari model kita (apakah weak atau strong dengan menggeser slider ke kiri atau ke kanan). Disini pada Regulation Typenya kami memilih Ridge (L2) dan strengthnya ke arah Strong sebagai bentuk defaultnya, setelah itu bisa close saja pop up logistic regression tsb.

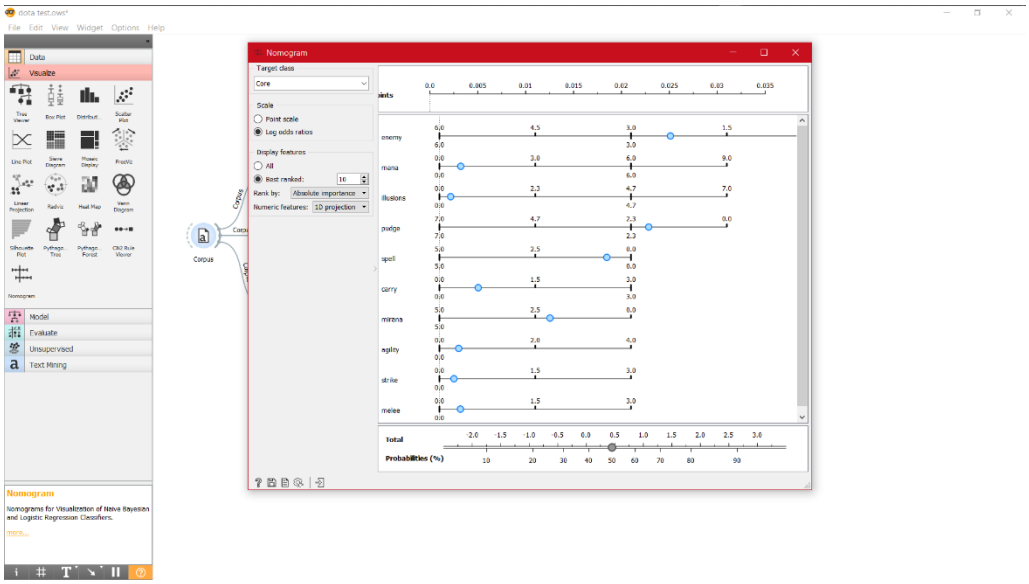
- **MENERAPKAN VISUALISASI CLASSIFIER DENGAN NOMOGRAM.**

Fungsi dari Nomogram itu untuk memvisualisasikan Classifier dari suatu model Logistic Segression. Widget Nomogram ini akan menampilkan Top Rank-10 kata yang paling penting dari Classifier dan Target Class dari **Dota Train**.

Untuk mencoba melakukan Nomogram pada software Orange, kita klik pada Widget/Add on di sebelah kiri pada bagian Visualize dan pilih Nomogram, bisa dengan cara drag dan drop ke canvas yang ada di sebelah kanan. Seperti gambar dibawah ini :



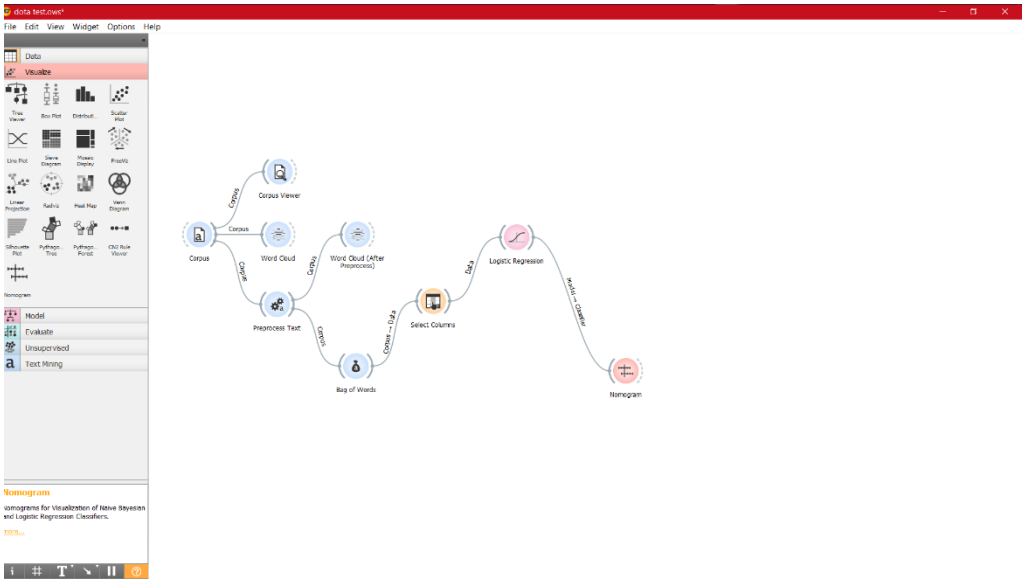
Setelah itu silahkan double click pada Nomogram maka tampilannya akan seperti dibawah ini :



Pada gambar diatas kita memilih Target Class yaitu “Core” dan kita bisa melihat bahwa yang berada pada posisi paling atas adalah kata yang paling berkontribusi untuk prediksinya. Contohnya seperti kata “enemy” dapat memberi tahu banyak kepada kita tentang makna dari dokumen tersebut. Jika kata “enemy” ini sering muncul di teks, dan menurut model Logistic Regression itu adalah deskripsi/penjelasan Tentang Role yang termasuk “Core” dalam Corpus **Dota Train**. Semakin ke kanan value-nya maka semakin besar untuk Nomogram-nya.

Sedangkan Pada bagian bawah ada penjelasan dari total prediksi atas kata yang paling berkontribusi dalam sebuah dokumen/corpus, dan Probabilitas kata atas prediksi tersebut sebesar 50%, tentunya ini dalam atau menurut model Logistic Regression ya.

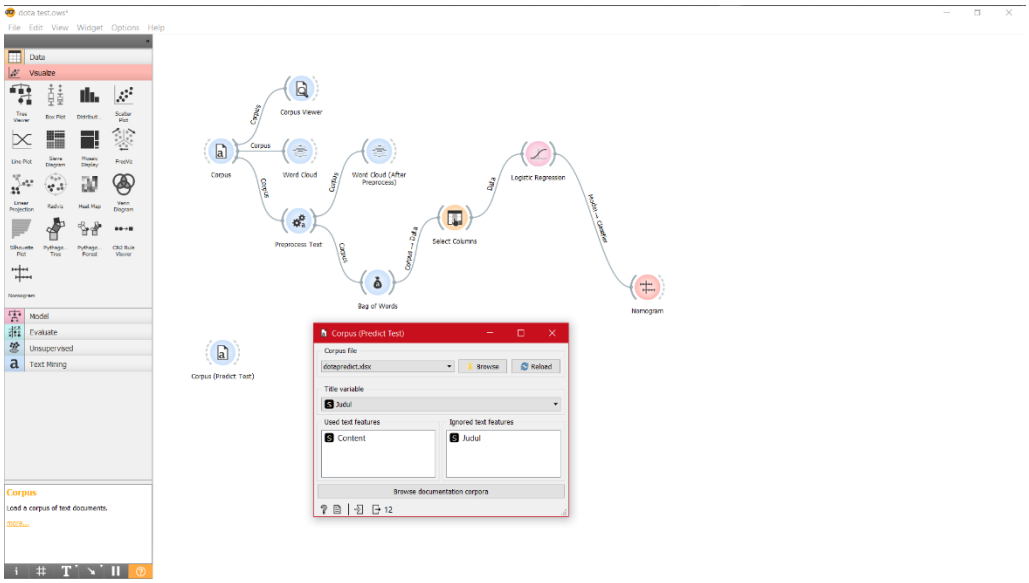
Tampilan Model Workflow kami sampai visualisasi dengan Nomogram.



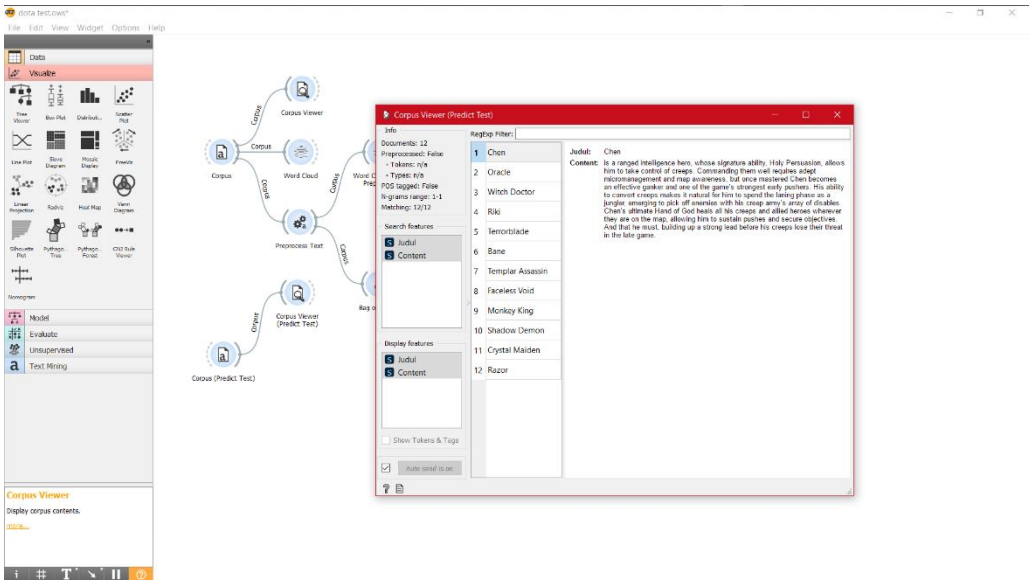
Sebelumnya kita pernah mengatakan sebagai tujuan awal bahwa kita ingin memprediksi data dari corpus dan kita tidak akan memprediksi sesuatu yang sudah kita tahu sebelumnya. Ini jelas merujuk ke dataset/corpus **Dota Train**. Nah, untuk melakukan itu semua dan mencoba untuk menguji performanya kita akan membuat Flow baru dibawah workflow dari Classifier kita sebelumnya

1. Pertama–tama kita buat corpus baru. Kemudian pilih corpus dari **Dota Predict Test**. Corpus ini sebelumnya sudah di jelaskan pada halaman awal, Corpus Dota Predict Test berisi 12 file dokumen, dan isinya mencakup Judul dan Content yang tentunya ini sebagai tujuan kami untuk membuat prediksi dan coba untuk mengklasifikasikan Role dari beberapa Hero yang ada di Dota 2 tersebut.

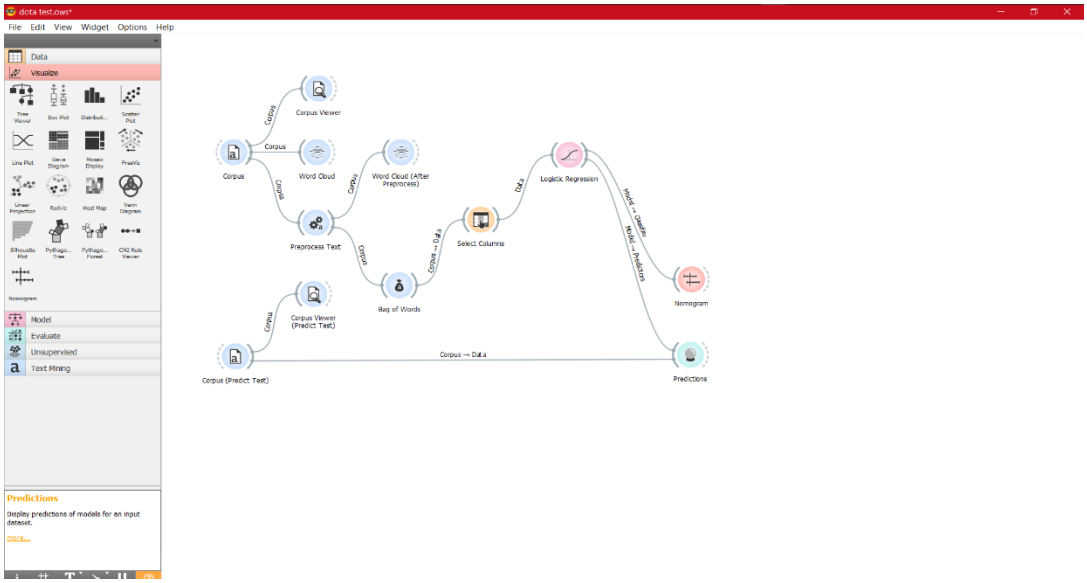
Untuk lebih jelasnya bisa Lihat tampilan dibawah ini :



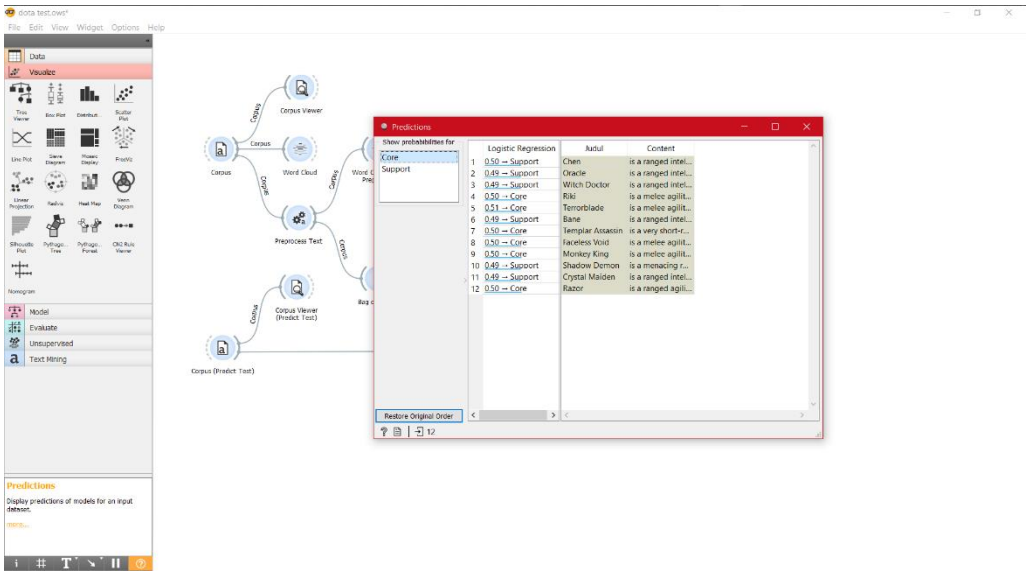
Setelah itu coba tambahkan Corpus Viewer untuk melihat dari isi dokumen Dota Predict tersebut. Pada Corpus Viewer disini terlihat **Dota Predict Test** ini Memiliki 12 dokumen seperti yang telah disebutkan diatas. Seperti terlihat pada gambar dibawah ini :



Selanjutnya disini kita akan mencoba untuk membuat prediksi. Caranya kita coba membuka widget/add on yang ada diseblah kiri dan arahkan pada bagian ***Evaluate*** lalu pilih yang namanya ***Prediction.*** Setelah itu coba hubungkan Corpus Dota Predict Test ke Prediction, dan hubungkan juga model dari Logistic Regression kita sebelumnya ke Prediction. Berikut seperti dibawah untuk tampilannya :

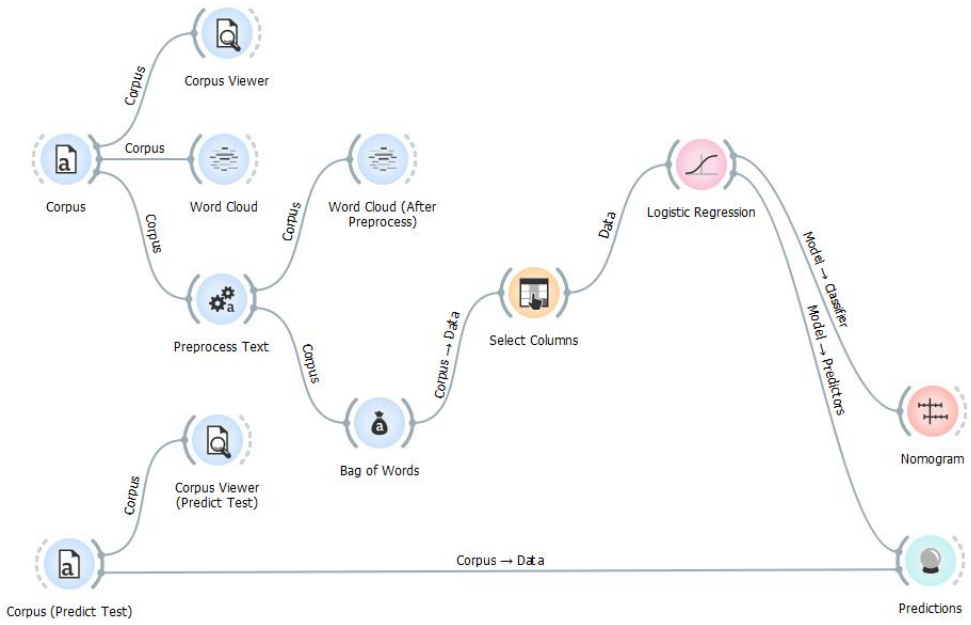


Setelah itu kita coba double-click pada Prediction untuk melihat hasil dari prediksi tersebut



Dan ternyata hasil dari prediction model ini memberi tahu kita jika dokumen corpus **Dota Predict** yang berjumlah 12 dokumen ini ketika diprediksi terutama yang berjudul “Chen” ini adalah bagian Role Model dari “Support”, lalu yang Berjudul “Riki” adalah Role Model dari “Core” dan sampai pada dokumen terakhir yang berjudul “Razor” adalah Role Model dari “Core”. Sepertinya benar dan terlihat menjanjikan bukan? Tetapi begitulah yang kita dapat dari model yang berhasil kita bangun menggunakan model *Logistic Regression* Khususnya dengan bantuan Prediction yang menggunakan oleh Software Text Mining itu. Meskipun Probabilitas yang dihasilkan pada Corpus diatas sebesar 50% tetapi bisa menghasilkan Prediksi yang pas dalam mengklasifikasikan Role Model dari Game Dota 2.

Model kita sudah jadi dengan tampilan full workflow seperti dibawah ini :



Terima Kasih