Implementasi Proses Data Science Pada Dataset Perusahaan MW Car Insurance Dengan Menggunakan Model Decision Tree, Logistic Regression, dan Gradient Boosting

William

Fakultas Teknik dan Informatika Universitas Multimedia Nusantara,

Tangerang, Indonesia

[william13@student.umn.ac.id](mailto:william13@student.umn.ac.id)

*Abstrak*— Di era industri 4.0, data memiliki peran yang sangat krusial dalam perkembangan teknologi dan implementasinya dalam kehidupan sehari-hari, salah satunya dalam sektor bisnis. Dengan menggunakan bantuan beberapa metode dan algoritma untuk menganalisa data dan mengolah data menjadi informasi yang berguna dapat digunakan menjadi dasar utama suatu perusahaan dalam pengambilan keputusan, salah satunya dalam sektor bisnis asuransi kendaraan. Dengan mengimplementasikan data mining dalam penelitian dan implementasi *Data Science* pada data cutomer, perusahaan bisa menganalisa *behavior* dan mendapatkan insight baru untuk menjalankan proses bisnis berdasarkan kebiasaan pelanggan sehingga proses klaim asuransi bisa menjadi optimal dan akurat. Objek penelitian yang dilakukan dalam perusahaan MW Car Insurance dibagi menjadi beberapa variabel utama yang berkaitan dengan proses klaim asuransi pada kendaraan, seperti pengalaman berkendara, umur kendaraan, tipe kendaraan, hingga data-data mengenai pelanggaran lalu linstas yang dilakukan. Dalam penelitian ini, saya akan membandingkan beberapa model yang digunakan dalam proses data science, yakni model Decision Tree, Logistic Regression, dan Gradient Boosting. Hasil dari penelitian ini diharapkan dapat membantu memberikan insight terhadap perusahaan MW Car Insurance agar memperkecil peluang suatu kendaraan melakukan klaim asuransi dengan memahami korelasi yang terbentuk antara variabel-variabel utama yang dianalisa.

Kata Kunci — Data Science, Decision Tree, Gradient Boosting, Data Mining, Logistic Regression, SAS Studio, Asuransi, Kendaraan

# I. LATAR BELAKANG & PEMAHAMAN BISNIS

## A. Latar Belakang

Dalam setiap transaksi pembelian kendaraan roda empat maupun roda 2, hampir setiap perusahaan menawarkan layanan produk asuransi yang ditujukan untuk melindungi kendaraan dari hal-hal yang tidak diinginkan, seperti kehilangan, rusak, dan berbagai maslaah lainnya. Alih-alih harus membayar dan merogok kocek sendiri untuk membayar kecelakaan mobil, mayoritas dari pemilik mobil maupun motor memilih untuk membayar premi asuransi terhadap kendaaran mereka masing-masing dengan berbagai jenis varian harga dan jenis proteksi yang di dapat.

Dengan hal ini, secara langsung pemilik kendaraan melakukan investasi dalam bentuk asuransi terhadap kendaraan mereka.

Dalam kasus spesifik pada asuransi kendaraan roda 4 (khususnya mobil), asuransi mobil secara efektif merupakan kontrak antara pemilik kendaraan dengan perusahaan asuransi yang preminya bisa dibayar berbarengan dengan cicilan perbulan kendaraan yang ditujukan untuk melindungi dari bentuk kerugian finansial yang di sebabkan oleh kendaraan tersebut. Informasi detail mengenai jenis tanggungan apa yang ditanggung dalam perusahaan asuransi tergantung dengan biaya premi dan perusahaan asuransi yang dipilih oleh pihak pemilik kendaraan.

Beberapa contoh perusahaan asuransi mobil yang cukup terkenal di Indonesia diantaranya yakni AXA Car Insurance, Alianz, Adira Insurance, Chubb General Insurance, MNC Insurance, dan perusahaan asuransi dari grup Sinarmas, yakni Sinarmas Car Insurance. Sistem asuransi kendaraan mobil di Indonesia sendiri biasanya dijadikan 1 oleh sales yang menawarkan beberapa pilihan dan biaya premi dari asuransi tersebut digabung menjadi 1 didalam total biaya cicilan mobil jika mengambil metode pembayaran secara mencicil dan apabila pemilik kendaraan melakukan purchasing mobil secara tunai, maka pemilik kendaraan bisa mendaftarkan asuransi kendaraan mereka secara pribadi ke perusahaan asuransi mobil atau melalui layanan konsultasi sales mobil tersebut.

Pengaplikasian dan proses klaim asuransi kendaraan di Indonesia sendiri masih kurang optimal dan transparan dan cenderung memakan waktu yang sangat lama dan tidak sebanding dengan kerusakan yang dialami. Pada umumnya, pembeli kendaraan kurang memperhatikan *Terms & Condition* dari polish asuransi kendaraan mereka saat melakukan pembelian di dealer showroom masing-masing sehingga sering terjadi miss komunikasi dengan p. Perusahaan MW Car Insurance menawarkan jasa asuransi kendaraan dengan polish yang terjangkau serta transparan sesuai dengan *Terms & Condition* yang terlampirkan sesuai dengan paket asuransi yang dipilih dan disertai dengan proses claim yang cenderung mudah sesuai dengan prosedur SOP klaim kendaraan customer.

## B. Pemahaman Bisnis

Dalam tahap pehamanan bisnis, perusahaan harus menentukan tujuan serta visi dan misi dari proses bisnis yang dilakukan. Perusahaan MW Car Insurance harus memahami konteks bisnis dan cakupan-cakupan yang sesuai dengan bidang bisnis yang dijalani. Dalam konteks perusahaan MW Car Insurance, perusahaan harus memahami objective dari perspektif customer (dalam konteks ini pemilik kendaraan mobil yang mendaftar asuransi) maupun dari perspektif perusahaan serta melakukan proses perekrutan tenaga kerja yang berkualitas dan memada untuk mendukung proses bisnis yang dilakukan.

Tujuan dari laporan ini yakni untuk menganalisa customer behavior / kebiasaan yang dilakukan customer pada saat melakukan klaim asuransi kendaraan berdasarkan dataset MW Car Insurance dan menganlisa faktor apa saja yang mempengaruhi besarnya peluang pemilik kendaraan dalam melakukan klaim asuransi kendaraan. Proses bisnis MW Car Insurance diharapkan akan berjalan secara opimal dengan memanfaatkan analisa dari project Data Science yang dilakukan dan perusahaan dapat mengetahui penyebab utama apa yang menyebabkan pengendara mobil melakukan klaim asuransi dan dapat mengurangi peluang untuk melalukan klaim asuransi mobil dengan memperbaiki faktor tersebut.

# II. TINJAUAN TEORITIS & METODOLOGI PENELITIAN

## A. Tinjauan Teoritis

1. ***Definisi, Kegunaan, serta keuntungan dari Algoritma Decision Tree***

*Decision tree* adalah salah satu algoritma yang digunakan untuk melakukan klasifikasi didalam data analysis. Decision Tree akan mengubah data menjadi *Decision Rules* dalam bentuk struktur pohon yang dimana setiap nodenya merepresentasikan atribut sedangkan batang merepresentasikan value dari atribut tersebut. Decision Tree sering juga disebut sebagai CART (Classification and Regression Tree).

*Decision Tree* pertama kali digunakan pada tahun 1966 oleh Institute of Computing Science Pozman University of Technology, Polandia yang digunakan untuk *human concept model project* dan terus berkembang hingga tahun 1974 ahli statistika asal UC Berkeley California, Charles Stone dan Leo Breiman mulai mengembangkan algoritma Decision Tree /CART dengan partnernya, Jerome Friedman dan Richard Olsen dari Stanford University. Mereka merilis algoritma CART ke kalangan publik pada tahun 1984 dalam bentuk software yang digunakan sebagai salah satu metode research data anaylsis.

Sejak saat itu, CART terus berkembang dan melahirkan beberapa tipe algoritma *Decision Tree* baru, seperrti ID3 (Iterative Dichotomiser 3) yang menggunakan *gain ratio* dan C4.5 yang menduduki peringkat pertama dalam 10 besar algoritma yang digunakan untuk melakukan data mining pada tahun 2008.

Keuntungan utama yang didapat jika menggunakan algoritma Decision tree yakni kemampuannya untuk melakukan *break down* terhadap data dan membuat tampilannya menjadi lebih sederhana. Algoritma ini juga berguna untuk menemukan hubungan yang tercipta antara beberapa variabel dalam dataset, seperti variabel input dan variabel target dalam perbandingan. *Decision Tree* menjelaskan opsi, resiko, tujuan serta manfaat dari setiap opsi yang ditawarkan. Berbeda dengan algoritma lainnya (Misalkan KNN), *Decision Tree* dapat mencegah terjadinya bias pribadi karena akan menghasilkan perspektif yang seimbang dari resiko yang dihitung. *Decision Tree* juga menyajikan infomasi dengan cara yang sederhana dan mudah dipahami sehingga dapat dianalisis dengan cepat dan membantu kita dalam mengambil keputusan penting secara efisien.

1. ***Definisi, Kegunaan, serta keuntungan dari Algoritma Logistic Regression***

*Logistic Regression* merupakan algoritma dalam data analyis yang digunakan untuk mengecek hubungan dan pengaruh anatara 2 atau lebih variabel. Hasil dari *Logistic Regression* biasanya digunakana sebagai referensi untuk menggambarkan konklusi serta kesimpulan apakah terdapat korelasi antara variabel yang dibandingkan atau tidak. Metode ini cocok digunakan untuk memprediksi values yang akan terjadi di masa depan, seperti pergerakan saham.

*Logistic Regression* sangat *Suittable* ketika kita sedang mencari hubungan antara variable independent dan variabel dependent. Ketika terdapat hubungan linear antara kedua variable tersebut, kita bisa mengecek multikolinearitas dari data, auto korelasi, dan juga kita bisa melihat alur distribusi data apakah terdistribusi secara normal/tidak.

Dalam algoritma Logistic Regression, kita harus menghitung probabilitias yang terletak diantara interval 0 dan 1, kemudian probabilitas tersebut dapat digunakan untuk mengklasifikasikan data. Misalnya jika probabilitas yang dihitung lebih besar dari 0,5 maka data tersebut termasuk dalam kelas A dan sebaliknya. Jika kurang dari 0,5 maka data tersebut termasuk ke dalam kelas B. Dalam uji test ini setidaknya ada kondisi yang harus dipenuhi, yakni setidaknya 2 variabel predictor dalam model linear dan test terakhir yakni RMSE untuk melihat presentase error dala prediksi yang dibuat.

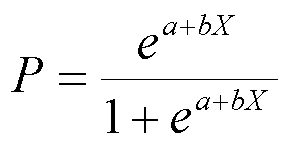


Figure Simple Logistic Regression Formula

## *Definisi, kegunaan, serta keuntungan dari Gradient Boosting*

Gradient Boosting merupakan salah satu teknik yang digunakan untuk membuat modelling dalam machine learning yang digunakan untuk klasifikasi masalah. Gradient boosting dirancang dan dikembangkan dengan pendekatan dan fundamental yang berbeda dengan metode tradisional, seperti Decision Tree dan juga Random Forest.

Kelebihan utama dari metode ini yakni kecepatan prediksi dan akurasi yang tinggi. Selain itu, metode gradient boosting juga cocok digunakan dalam dataset yang memiliki cakupan yang besar dan kompleks. Gradient Boosting digunakan untuk mencari hubungan non linear antara model target dan features dan sangat bisa digunakan untuk mengatasi masalah *missing values,* nilai cardinal yang tinggi, dan juga outlier tanpa ada metode khusus.

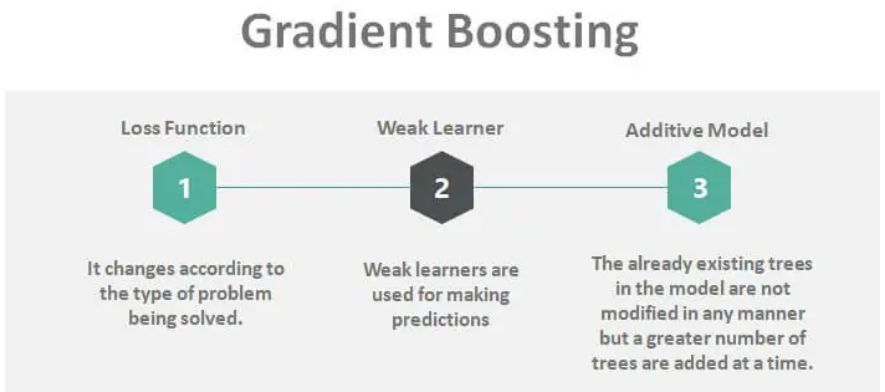


Figure Cara kerja Gradient Boosting

## B. Metode Penelitian

Metode penelitian yang digunakan dalam artikel jurnal ini yaitu metode penelitian kuantitatif yang terdiri dari angka dan hasil akurasi dari pengetesan model dengan tujuan untuk menentukan hubungan antara satu variabel (independen) dan variabel lainnya (dependen) serta membandingkan tingkat hasil akurasi antar model dan menentukan model mana yang memiiki tingkat akurasi terbaik

Data yang digunakan dalam penelitian ini menggunakan data deskriptif dan kuantitatif yang berisi tentang variabel-variabel utama serta angka / nilai yang mempengaruhi pemilik kendaraan untuk melakuka klaim asuransi mobil mereka.

Penelitian kali ini menggunakan dataset yang didapat dari salah satu website penyedia dataset yang bisa diakses secara gratis, yakni melalui website kaggle.com. penulis sudah memastikan dataset yang digunakan dalam penelitian kali ini tidak memiliki data NULL dan variabel yang terdapat dalam dataset memiliki hubungan yang relevan dengan topik yang dibahas.

## C. Objek Penelitian

Objek Penelitian yang digunakan yakni menggunakan dataset dari perusahaan MW Car Insurance yang berisikan mengenai informasi-informasi dan riwayat seputar kendaraan dan pengalaman berkendara masing-masing pemilik mobil. Selain itu, status sosial dan status keluarga pemilik kendaraan juga merupakan salah satu objek yang akan di teliti dalam penelitian kali ini.

## D. Pemahaman Data

Proses *data understanding* merupakan proses pengenalan dan pemahaman dataset perusahaan yang akan digunakan dalam proses data science. Proses utama dalam data understanding secara umum dibagi menjadi 3 proses utama, yakni pendeskripsian data seperti menghitung jumlah rows dan berapa banyak variabel yang terdapat di dalam dataset yang akan digunakan.

Didalam dataset perusahaan MW Car Insurance sendiri, terdapat 19 kolom dan 987 baris. Terdapat 16 variabel utama yakni yang akan di analisa, yakni ID, Age, Gender, Race, Driving Experience yang dikategorikan berdasarkan tahun berkendara, Education, Income, Vehicle Ownership, Vehicle Year, Speeding, Past Accident, Outcome Income, dan beberapa status sosial seperti status pernikahan (Married), Children, Alamat berupa postal code, dan tipe spesifik kendaraan (Vehicle Type). Masing-masing variabel memiliki keterkaitan dengan proses klaim asuransu kendaraaan. Berikut merupakan tampilan secara keseluruhan dataset:

Sebuah gambar berisi meja

Description automatically generated

Figure 3 Tampilan Secara Keseluruhan Dataset MW Car Insurance

# III. PERSIAPAN DAN EKSPLORASI DATA

## A. Persiapan Data

Tahap persiapan data mencakup proses untuk mengubah dataset menjadi bentuk tabel proporsional yang akan digunakan untuk proses data science selanjutnya, yakni tahap eksplorasi data dan juga data modelling. Saya menggunakan fitur Prepare Data dalam SAS Drive untuk melakukan proses Standardize. Dengan melakukan Standardize dalam tahap persiapan data, kita bisa menghasilkan analisa yang akurat karena menggunakan data dengan format yang sesuai, proporsional dan konsisten sehingga dapat menghasilkan keputusan yang akurat dalam proses bisnis MW Car Insurance. Berikut merupakan tahap persiapan data dalam SAS Drive dataset MW Car Insurance:

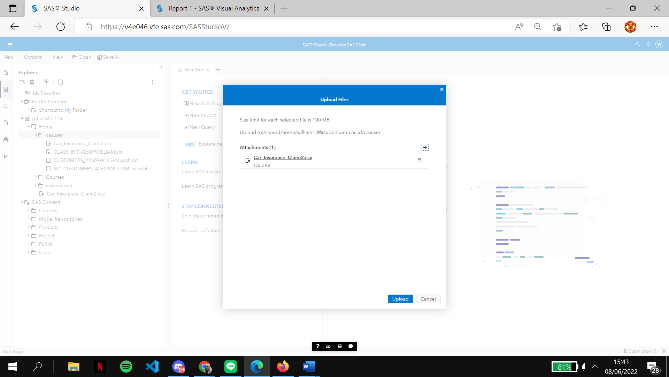


Figure 4 Import Car Insurance CSV

Sebuah gambar berisi meja

Description automatically generated

Figure 5 Memilih Data

Sebuah gambar berisi teks

Description automatically generated

Figure 6 Standardize Data (Postal Code)

Sebuah gambar berisi teks

Description automatically generated

Figure Standardize GENDER

Sebuah gambar berisi teks

Description automatically generated

Figure Standardize EDUCATION

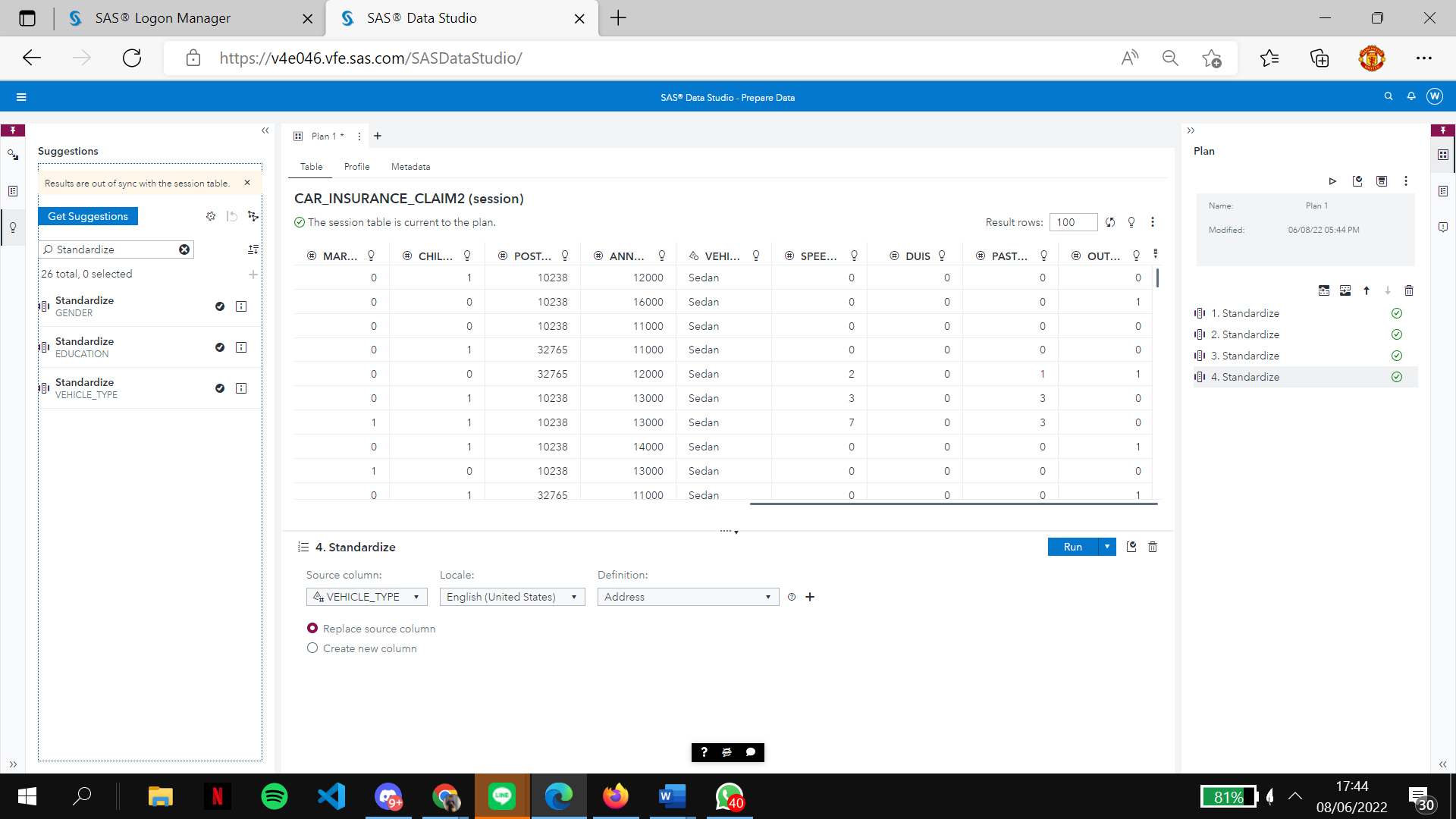


Figure Standardize VEHICLE\_TYPE

## B. Eksplorasi Data *(Exploratory Data Analysis)*

Setelah Tahap preparasi data selesai, langkah selanjutnya dalam mengimplementasikan proses *Data Science* yakni melakukan eksplorasi data. Pada proses ini data scientist menentukan data mana yang perlu dilakukan stukturisasi ulang dan cleansing jika diperlukan. Saya mengimplementasikan *Exploratory Data Analysis* dalam proses ini untuk mengetahui statisika deskriptif dan numerik pada masing-masing variabel yang digunakan dalam data items. Melalui tampilan pada Data Pane dalam SAS Visual Analytics, kita bisa mengcreate new data baru serta menentukan berbagai metode seperti hieararchy, effect, sampai dengan partition. Berikut merupakah proses *Exploratory Data Analysis* yang digunakan dalam menggunakan beberapa variabel utama dalam dataset perusahaan MW Car Insurance:

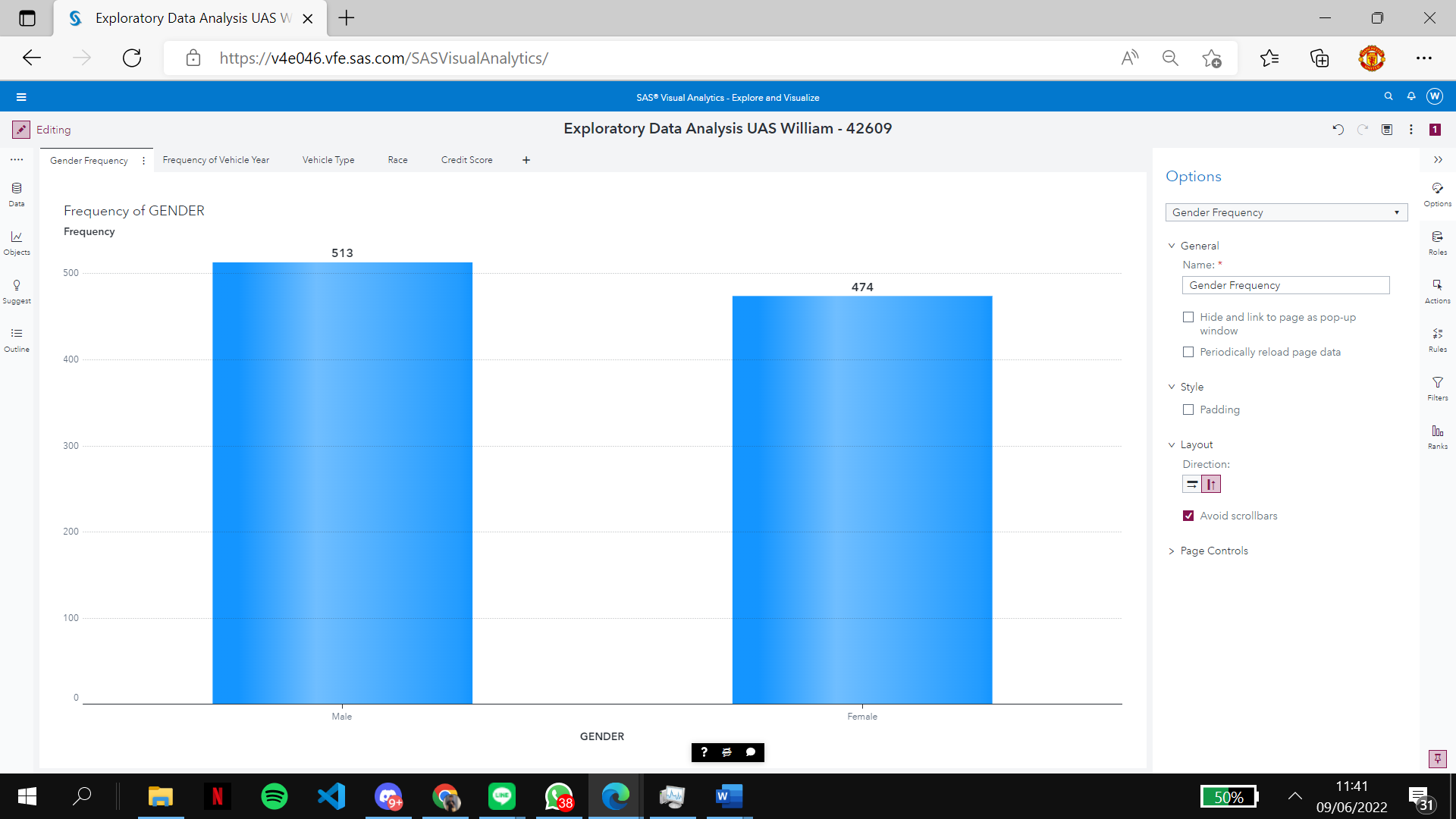


Figure Exploratory Data Analysis (Frequency of Gender)

Berdasarkan hasil visualisasi yang ditampilkan dari figure 10, dapat disimpulkan bahwa dari total 987 responden, jumlah responden pria dalam melakukan klaim asuransi lebih banyak dibandingkan dengan wanita meskipun memiliki selisih yang tipis (513 untuk pria dan 474 untuk wanita).

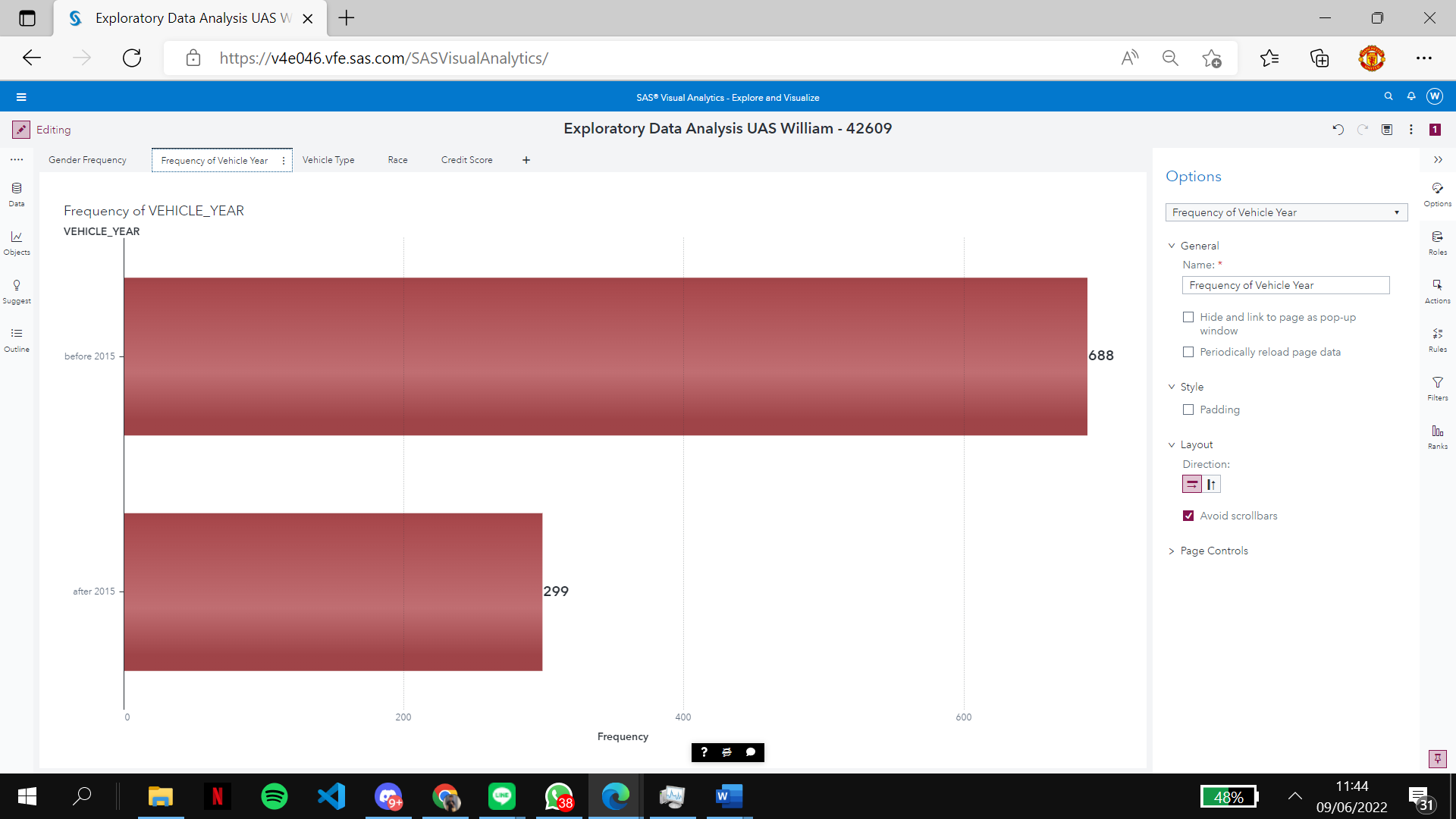


Figure Exploratory Data Analysis (Frequency of Vehicle Year)

Berdasarkan hasil visualisasi yang ditampilkan dalam figure 11, dapat disimpulkan bahwa frekuensi kendaraan produksi sebelum tahun 2015 merupakan jenis kendaraan yang paling banyak melakukan asuransi. Terdapat 688 kendaraan produksi sebelum tahun 2015 yang melakukan klaim dan 299 kendaraan setelah tahun 2015 yang melakukan klaim asuransi. Hal ini berbanding lurus dengan hipotesa mengingat semakin tua umur kendaraan, maka semakin besar peluang untuk mengganti parts / servis kendaraan.

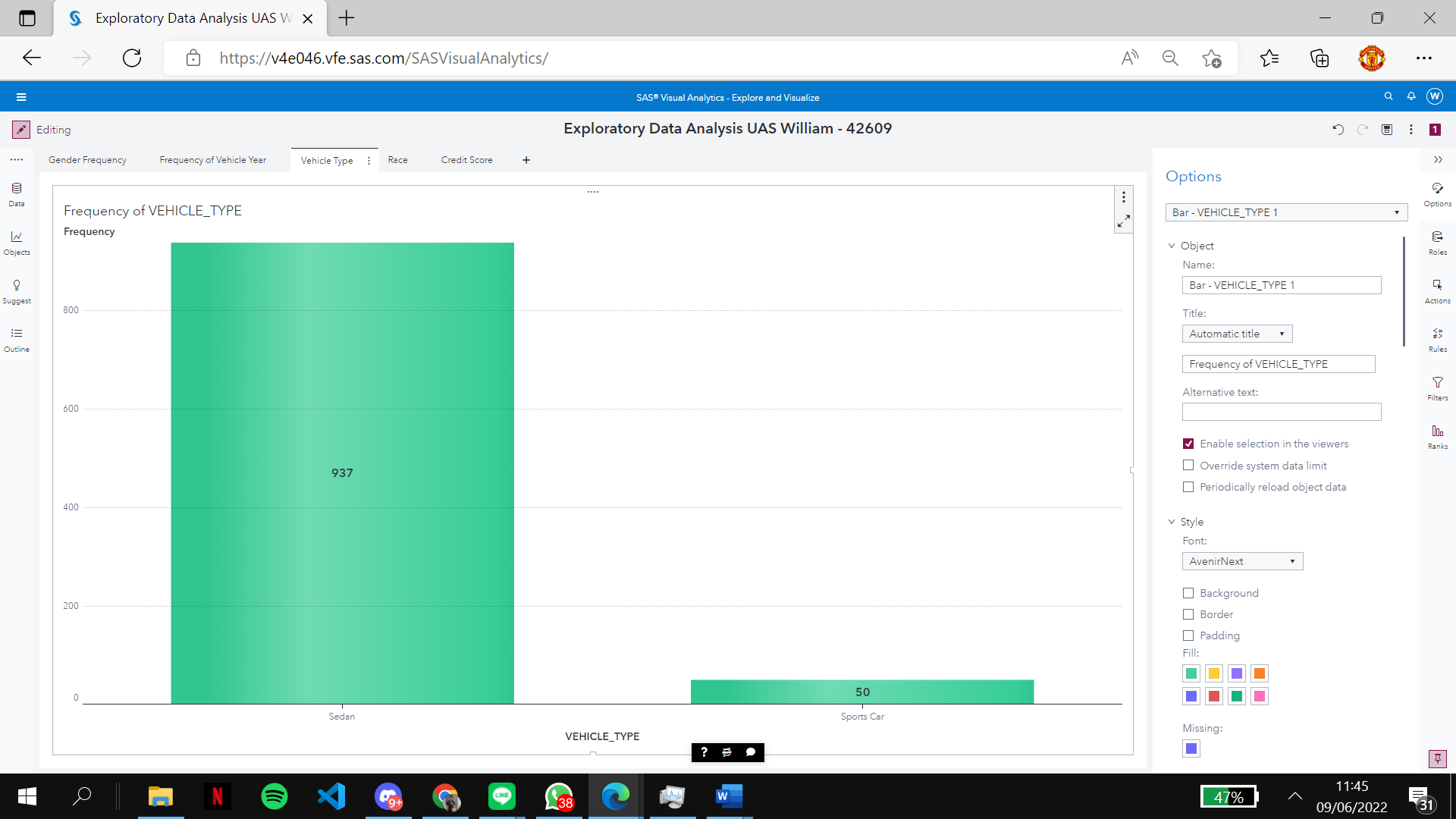


Figure Exploratory Data Analysis (Frequency of Vehicle Type)

Berdasarkan hasil visualisasi yang ditampilkan dari figure12, dapat disimpulkan bahwa jumlah tipe mobil sedan mendominasi jumlah responden yang melakukan klaim asuransi dengan jumlah 937 kendaraan, sedangkan mobil sports dengan jumlah 50 kendaraan. Hal ini berbanding lurus dengan hipotesa mengingat jumlah penjualan kendaraan sedan yang lebih banyak dibandingkan dengan mobil sport.

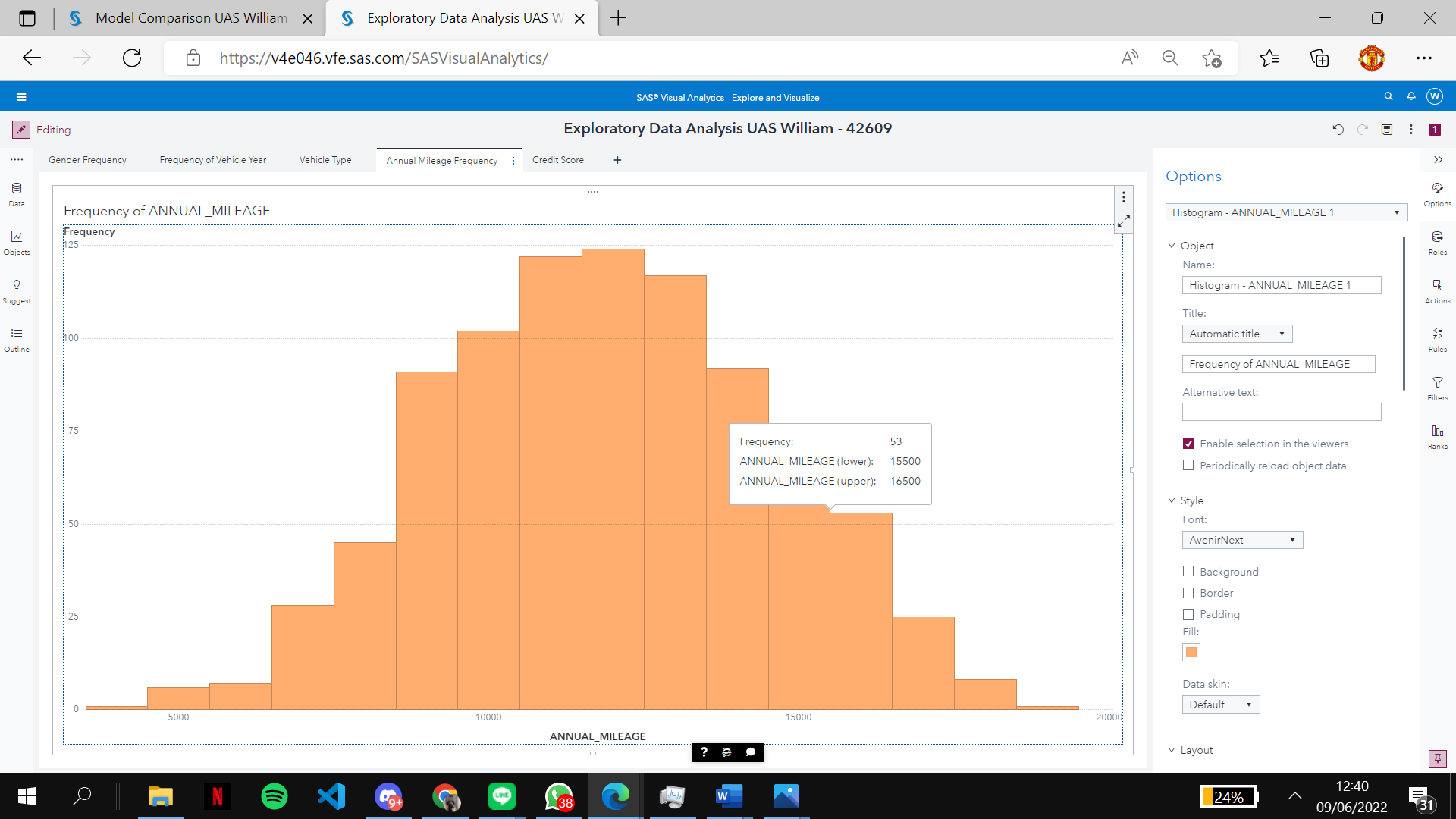


Figure Exploratory Data Analysis (Frequenct of Annual Mileage)

Berdasarkan hasil visualisasi yang didapat dari figure 13 yang memvisualisasikan Frekuensi dari ANNUAL MILEAGE dalam bentuk histogram, dapat disimpulkan bahwa jarak tempuh kendaraan antara 11.500-12.500 miles memiliki jumlah responden yang paling banyak, yakni 124 responden sehingga dapat disimpulkan mayoritas pemilik kendaraan yang melakukan klaim asuransi memiliki kendaraan dengan jarak tempuh yang cukup banyak.

# 

Figure Exploratory Data Analysis (Frequency of Credit Score)

Berdasarkan hasil visualisasi yang didapat dari figure 14 yang memvisualisasikan Frekuensi dari CREDIT\_SCORE dalam bentuk histogram, dapat disimpulkan bahwa nilai credit score antara 0.525-0.575 memiliki jumlah responden yang paling banyak, yakni 119 responden sehingga dapat disimpulkan mayoritas pemilik kendaraan yang melakukan klaim asuransi memiliki nilai credit score dalam skala sedang.

# IV. PEMODELAN DATA

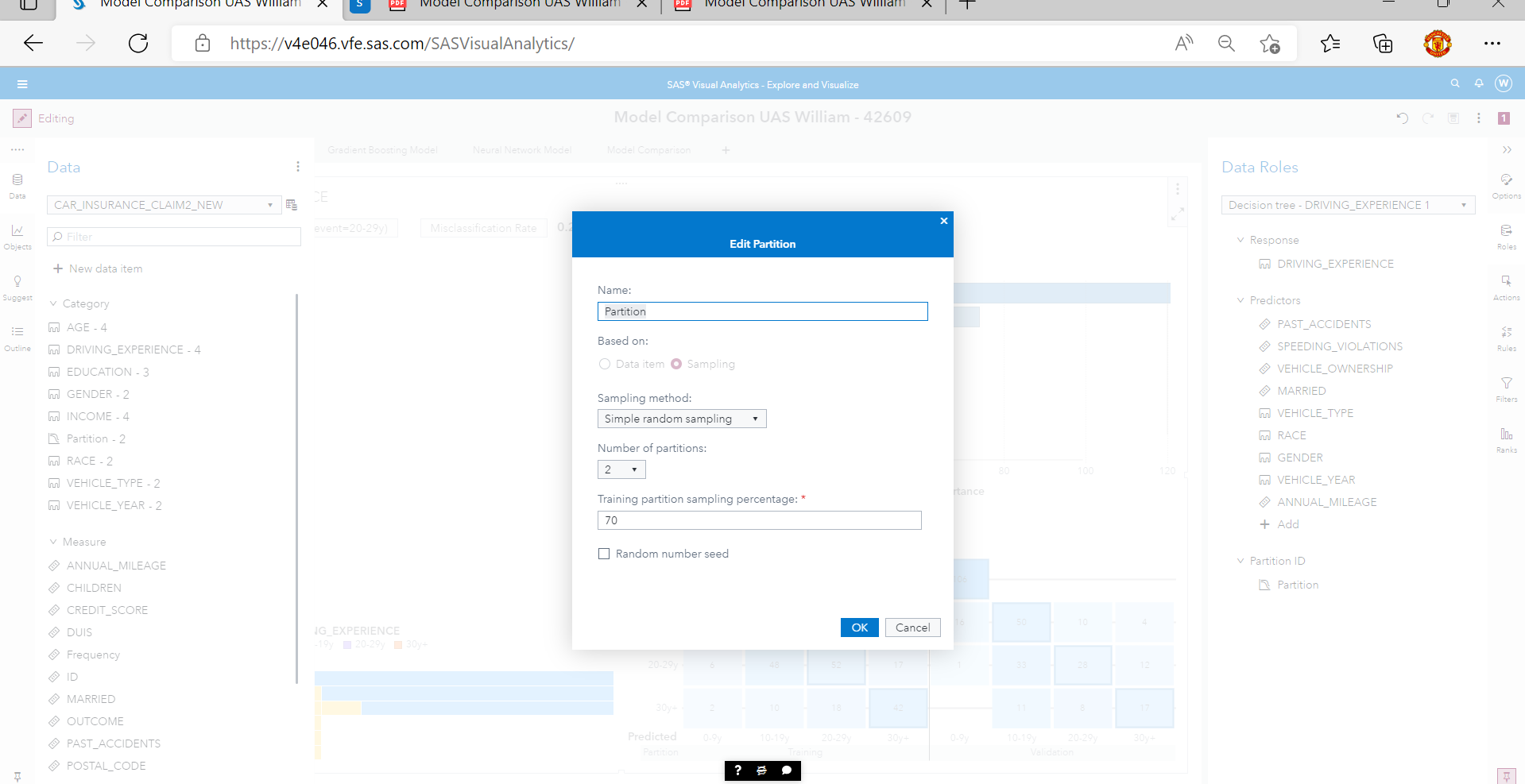


Figure Membuat Partition

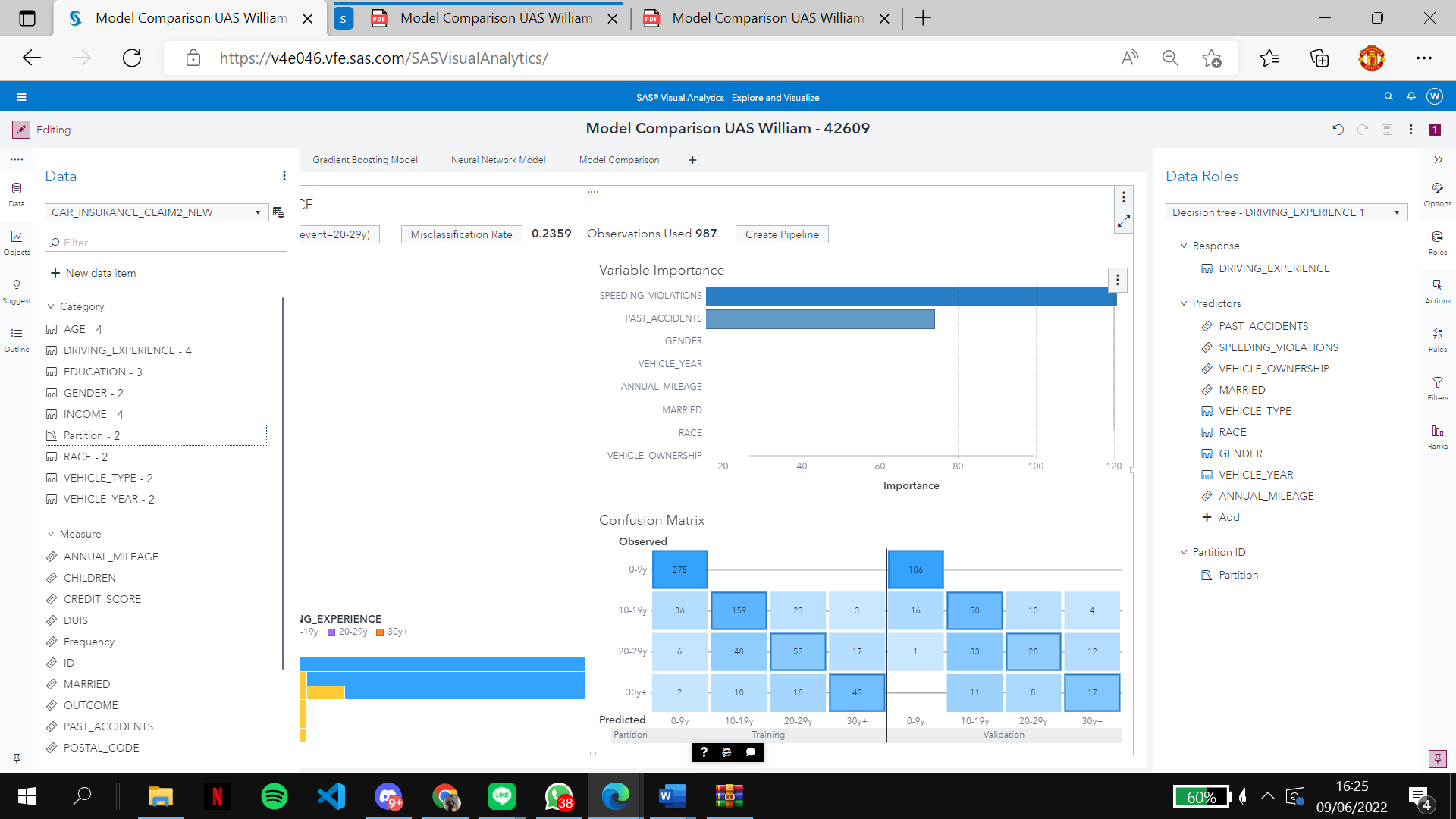


Figure Menentukan Data Roles

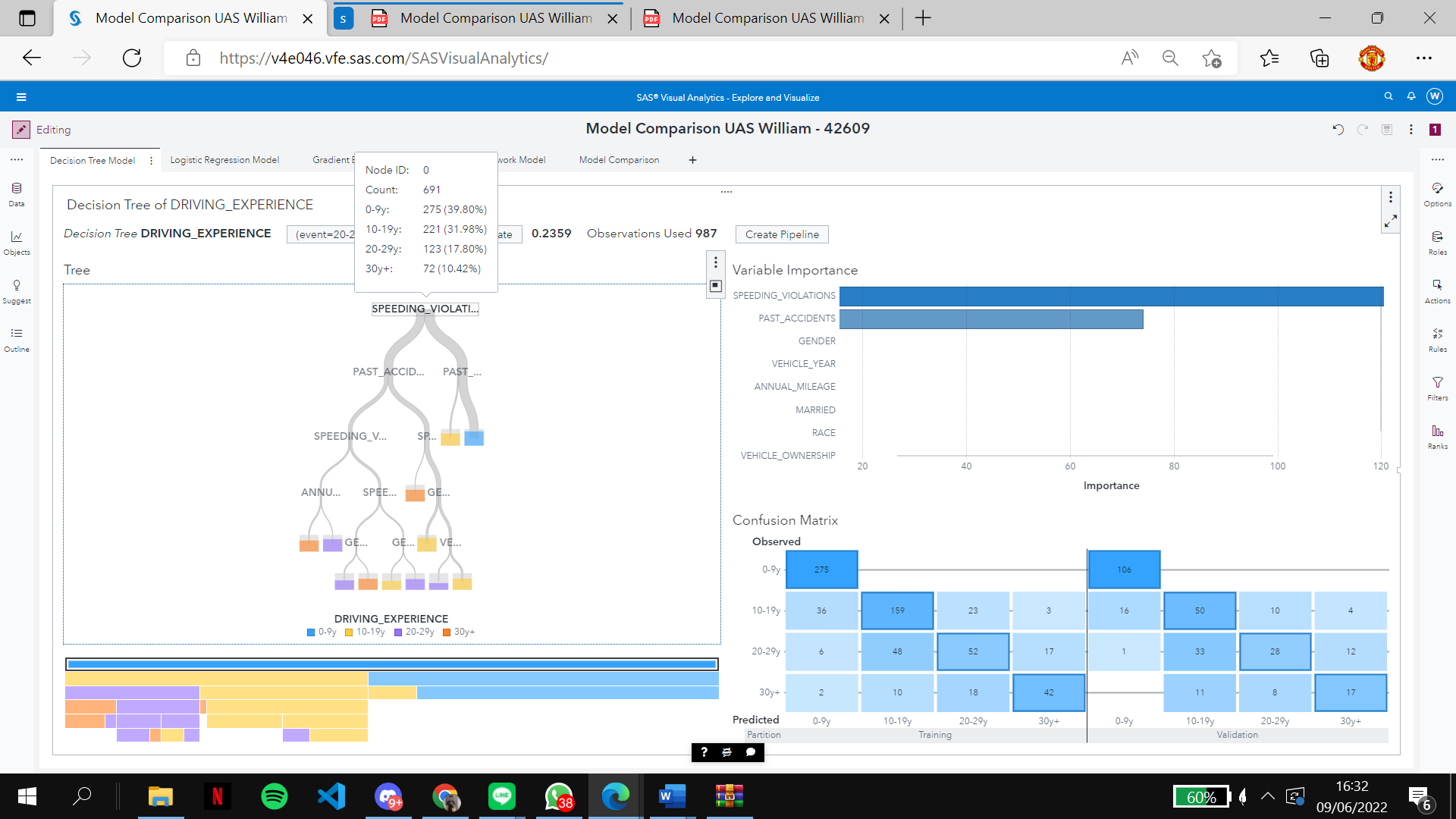


Figure Decision Tree Model

**A. Decision Tree Model**

Figure 15 menunjukan hasil visualisasi **Decision Tree Model***.* Decision Tree merupakan model yang paling umum dan cukup mudah dimengerti. Saya menggunakan beberapa 9 predictors seperti yang dilampirkan dalam figure 16 dan menggunakan DRIVING\_EXPERIENCE sebagai response dan Partition yang sudah di atur menggunakan partition sampling percentage sebesar 70% seperti yang terlampir dalam figure 15.

Dari hasil visualisasi model, didapatkan angka Training Misclassification Rate sebesar 0.2359 yang menunjukan tingkat kesalahan dalam memprediksi subjek sebesar 23,59% sehinnga memiliki akurasi sebesar 86,41%. Pada tampilan model Decision tree, terjadi pembagian pertama yang membagikan variabel SPEEDING\_VIOLATION menjadi 4 kelompok umur, yakni 0-9 dengan total 275 responden yang menjadikan kelompol umur terbanyak yang mengalami pelanggaran kecepatan (39,8 %) diikuti dengan kelompok umur 10-19 dengan total 221 responden (31,98%) lalu diikuti dengan kelompok umur 20-29 tahun dengan total 123 responden (17,8%) dan terakhir yaitu kelompok umur dengan 30+ year experience dengan total 72 responden (10.42%)

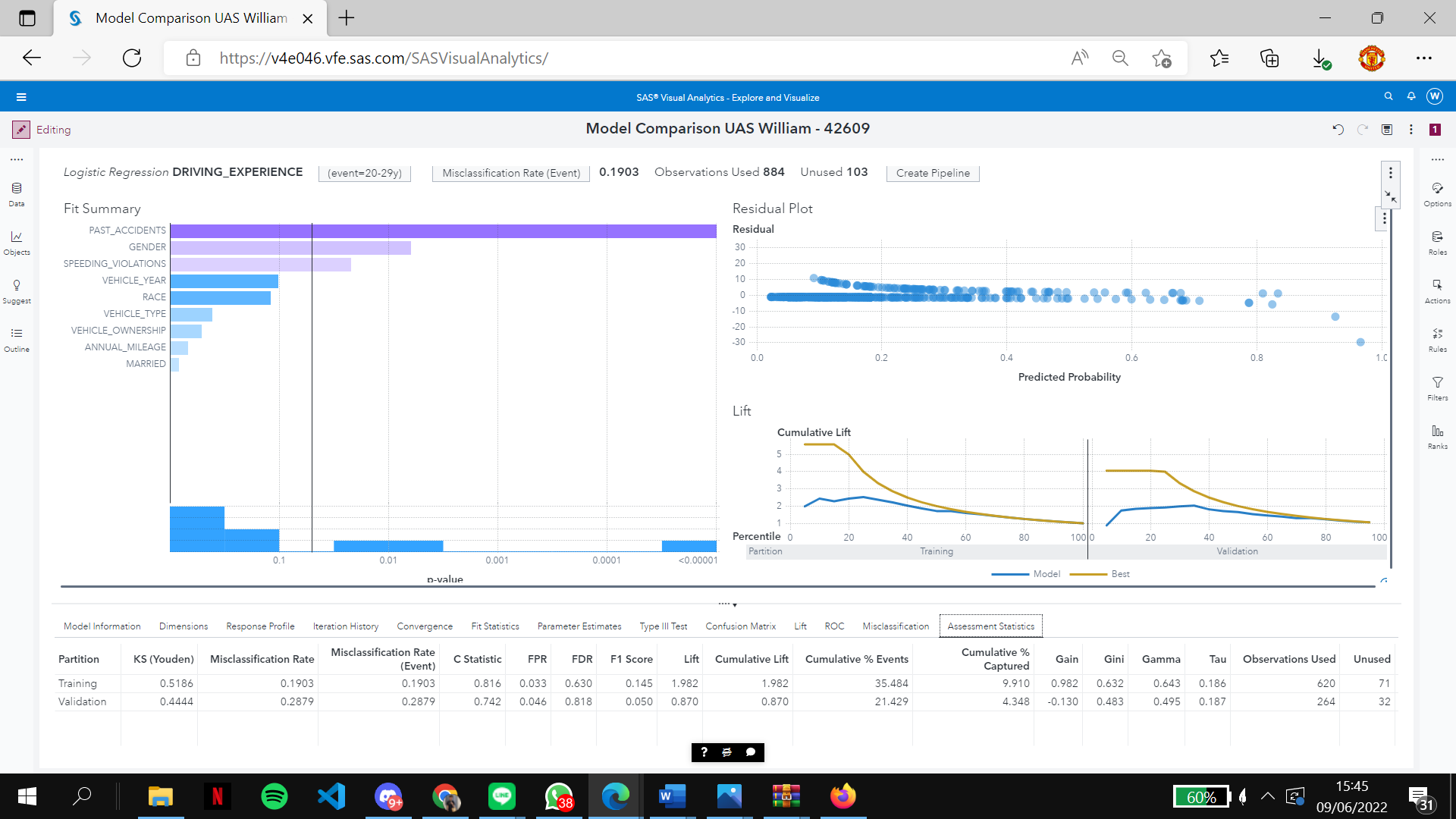


Figure Logistic Regression Model

## B. Logistic Regression Model

Figure 18 menunjukan hasil **visualisasi Logistic Regression Model**. Sama seperti dalam decisiont tree, saya menggunakan partisi dan Data roles yang sama dengan Decision Tree. Dari hasil visualisasi model, diperoleh angka Training Misclassification Rate (Event) sebesar 0.1903. yang menunjukan tingkat kesalahan dalam memprediksi subjek sebesar 19,03% sehinnga memiliki akurasi sebesar 80,97%. Dari hasil visualisasi juga menunjuka terdapat 103 unused / missing values dari predictors yang dipilih.

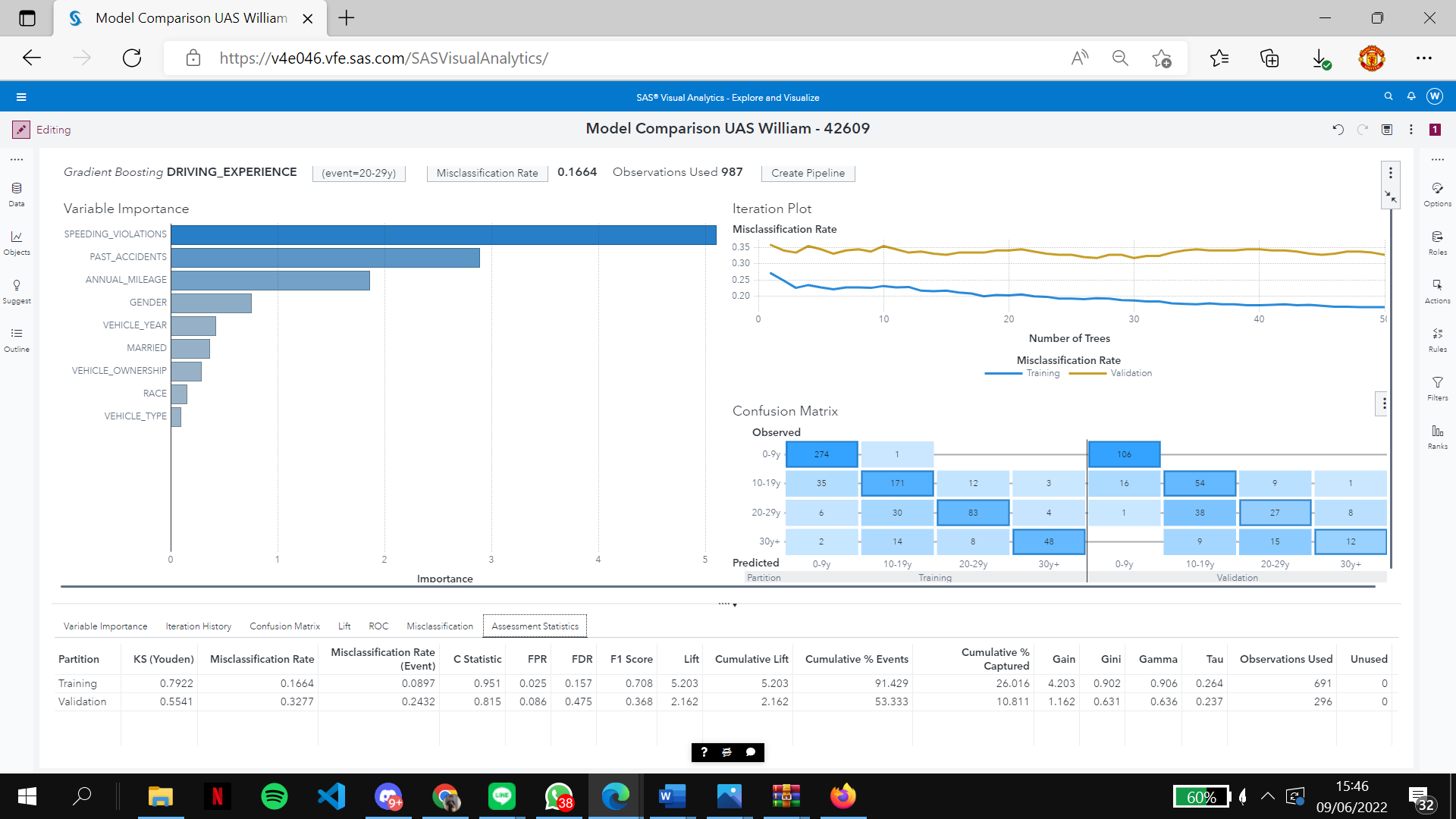


Figure Gradient Boosting Model

## C. Gradient Boosting Model

Figure 19 menunjukan hasil visualisasi **Gradient Boosting Model**. Gradient Boosting Model merupakan model yang kompleks dan menggabungkan teknik yang terdapat dalam deision tree dengan predictive model. Sama seperti dalam Decision Tree dan Logistic Regression, saya menggunakan partisi dan Data roles yang sama dengan Decision Tree. Dari hasil visualisasi model, diperoleh angka Training Misclassification Rate sebesar 0.1644. yang menunjukan tingkat kesalahan dalam memprediksi subjek sebesar 16,44% sehinnga memiliki akurasi sebesar 83,56%.

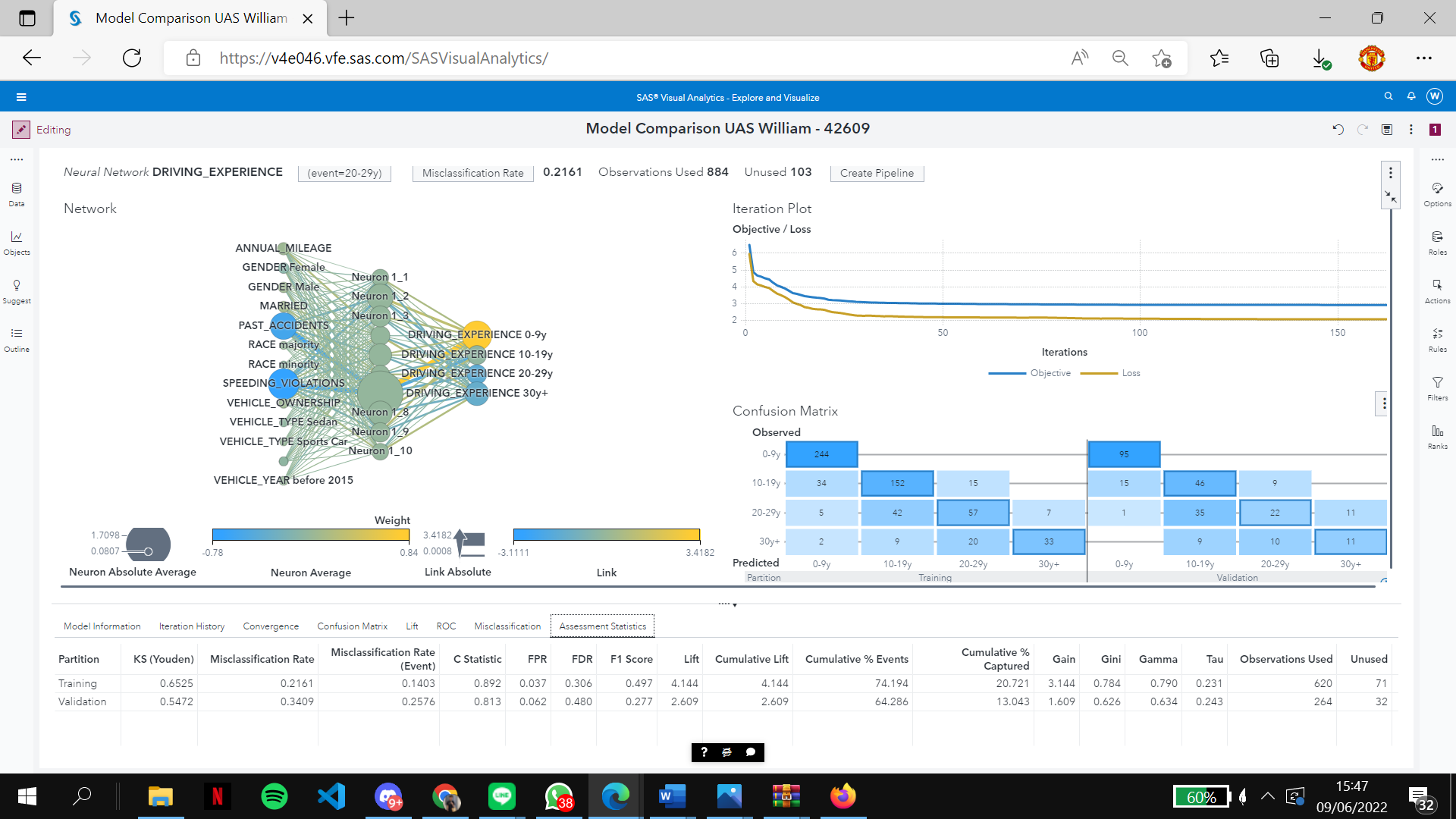


Figure Neural Network Model

## D. Neural Network Model

Figure 20 menunjukan hasil visualisasi **Neural Network Model**. Neural Network Model merupakan model yang menghasilkan visualisasi paling rumit dan menunjukan kumpulan *nested Regressions*  untuk mendapatkan hasil variabel. Sama seperti dalam Decision Tree ,Logistic Regression dan Gradient Boosting, saya menggunakan partisi dan Data roles yang sama dengan Decision Tree. Dari hasil visualisasi model, diperoleh angka Training Misclassification Rate sebesar 0.2161. yang menunjukan tingkat kesalahan dalam memprediksi subjek sebesar 21,61% sehinnga memiliki akurasi sebesar 78,39%.



Figure Model Comparison

## E. Determine the Best Model

Figure 21 menunjukan grafik perbandingan antara keempat model sebelumnya dengan menggunakan DRIVING\_EXPERIENCE sebagai variabel utama. Pada tampilan sebelah kiri menunjukan fit statistic Training nilai Misclassification Rate (Event) dari keempat model yang sudah dijelaskan di point sebelumnya, sedangkan pada tampilan sebelah kanan menunjukan Cumulative Lift dari data Training.

# V. KESIMPULAN

Masalah utama yang dihadapi oleh perusahaan asuransi kendaraan mobil umumnya berupa harga premi asuransi yang disama ratakan dan transparansi mengenai info detail peristiwa kecelakaan yang membuat pemilik kendaraan tersebut melakukan klaim asuransi. Berdasarkan hasil perbandingan modelling enggunakan Decision Tree, Logistic Regression, Gradient Boosting dan Neural Network Model, dapat disimpulkan bahwa Gradient Boosting merupakan model terbaik dengan nilai Misclassification Rate (Event) terendah dibanding dengan nilai model lainnya. Hasil analisa dari keempat model tersebut diharapkan dapat digunakan oleh MW Car Insurance untuk menyelesaikan dan menyesuaikan premi pembayaran asuransi sesuai dengan faktor-faktor utama penyebab pemilik kendaraan untuk melakukan klaim asuransi, seperti riwayat pelanggaran lalu lintas, sampai dengan pengalaman berkendara.

# UCAPAN TERIMA KASIH

Saya mengucapkan terimakasih sebesar-besarnya kepada Bapak Iwan Prasetiawan, S.Kom., M.M., sebagai dosen pembimbing mata kuliah IS 429 Big Data Analytics kelas A program studi Sistem Informasi Universitas Multimedia Nusantara yang telah membimbing saya selama berlangsungnya kelas 1 semester ini dan membantu mendiskusikan serta memberikan arahan mengenai tugas akhir mata kuliah IS 429 Big Data Analytics.

# DAFTAR PUSTAKA

Basavarajaiah D.M, B. N. (November 2020). Random Forest and Concept of Decision Tree Model. *Design of Experiments and Advanced Statistical Techniques in Clinical Research*, 133-156.

Gregor Stiglic, S. K. (2012). Comprehensive Decision Tree Models in Bioinformati. *Journal.pone.0033812 PubMed*.

Maciej Rzychoń, A. Ż. (May 2020). An Interpretable Extreme Gradient Boosting Model to Predict Ash Fusion Temperatures . *MDPI Journal Minerals*, 10(6):487.

Nemanja Milanovi´c, M. M. (2020). An Acceptance Approach for Novel Technologies in Car Insurance. *Faculty of Organizational Sciences, University of Belgrade, 11000 Belgrade, Serbia Sustainability*.

Rasheed, N. M. (2021). Diagnosis In The Ordinal Logistic Regression Model. *Turkish Journal of Compuer and Mathematics Education*, Vol 12 No.10 5524-5532.

Theodoros Petrakis, A. K. (May 20222). Neural Network Model for Greenhouse Microclimate Predictions. *Journal of Agriculture MDPI 12(6)*, 12.780.

Tzameret Rubin, T. H. (2022). Big data and data ownership rights: The case of car insurance. *Journal of Information Technology Teaching Cases*, Vol.0(0) 1-6.