A logo for a science company

Description automatically generated

**Ardian - 2106638173**

**Data Mining**

**Week 2 CRISP-DM Case Study**

**Case:**

Dikarenakan ChitChatGPT sering membuat joke yang kasar, project ini bertujuan untuk menghasilkan model machine learning yang dapat mengidentifikasi joke yang kasar. Dengan begitu, ChitChatGPT tidak lagi menghasilkaq jokes yang kasar.

**Answer:**

**BUSINESS UNDERSTANDING**

Dikarenakan ChitChatGPT sering membuat joke yang kasar, project ini bertujuan untuk menghasilkan model machine learning yang dapat mengidentifikasi joke yang kasar. Dengan begitu, ChitChatGPT tidak lagi menghasilkaq jokes yang kasar.

**DATA UNDERSTANDING**

Data yang digunakan untuk mentapai tujuan project ini adalah data teks, yang di mana teks tersebut abdalah teks jokes. Data teks jokes ini terdiri dari dua jenis, yaitu teks jokes yang kasar dan teks jokes yang tidak kasar. Data teks ini kemudian dipastikan kualitasnya, misal memastikan jenis jokesnya sesuai dengan label-nya (kasar atau tidak kasar), dsb.

**DATA PREPARATION**

Data teks kemudian diproses lebih lanjut melalui berbagai macam proses seleksi untuk membersihkan data data anomali, dan kemudian dibersihkan dengan proses text mining agar data tersebut siap untuk tahap modelling.

**MODELING**

Dengan data yang sudah siap, akan dilakukan pembuatan model yang membedakan jokes kasar dan tidak kasar. Pada tahap ini akan diseleksi algoritma atarpun arsitektur model yang terbaik, yang kemudian juga akan di-tune lebih lanjut untuk mendapatkan hasil model yang maksimal.

**EVALUATION**

Pada tahap ini, model dievaluasi menggunakan metrik yang sesuai. Evaluasi dilakukan untuk dengan tujuan memastikan bahwa model dapat mengidentifikasi joke yang kasar dengan baik dan juga tanpa mengorbankan performa pada joke yang tidak kasar. Setelah itu, dilakukan analisis lebih lanjut terhadap hasil evaluasi untuk memahami kekuatan dan kelemahan model serta menentukan apakah model tersebut sudah siap untuk dideploy.

**DEPLOYMENT**

Setelah model dianggap memenuhi kriteria yang ditetapkan, model tersebut dapat dideploy ke ChitChatGPT. Setelah itu, perlu juga dilakukan pemantauan dan pemeliharaan terhadap model yang sudah dideploy untuk memastikan performanya tetap optimal seiring berjalannya waktu. Tidak lupa juga untuk melakukan report terhadap hasil deploy dan pemantuan performa untuk dilakukan review lebih lanjut.