**Communications Skills:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Data to be Extracted*** | ***Features for Model*** | ***Preparation*** | ***Results - Weight Loss*** | ***Results - Abdominal Pain*** |
| Introduces self by full name  “Hi my name is Kai Jones” | * Greeting words (hello, hi, etc.) * “I am”, “My name is” “I’m” * Proper nouns via Google NLP * Location in conversation | * Done * Done * Done * Done |  |  |
| States student status  “I’m a medical student” | * “Medical student”, “student”, “I’m”, “I am” * Location in conversation |  |  |  |
| States purpose of encounter  “I’m just going to do a history and physical” | * “history”, “physical”, “I’m here to” * Location in conversation |  |  |  |
| **Open-Ended Question**  “Tell me more about the pain” | * Key words “what, how, tell me” * Response length |  |  |  |
| Interruptions | * Time between utterances * Audio data |  |  |  |
| “Allows for silence” | * Time between utterances * Audio data |  |  |  |
| Body Language and Non-verbal skills | * Video data   We will be able to get body language scores from patient + the video data and thus could predict the score based on the raw video footage later on |  |  |  |
| History was obtained in an organized fashion | * Plotting content-type over time * Topic modeling * Compare as a cohort |  |  |  |
| **Transition statements**  “I’m going to ask you some questions about your habits now” | * Key words (“I’m going to”, “I would like to”) * Use of “I” or “we” by physician |  |  |  |
| “Follow up questions are directed by patient answers” |  |  |  |  |
| **Repeated questions**  “Do you have a fever?”  1 min later  “Any fever?” | * Using sentence vectors find highly similar questions |  |  |  |
| Summarizes what has been said  “So far you’ve told me that you’ve lost some weight, have been having diarrhea… | * Summary words “summarize”, “so you’ve” * Sentence length * # of topics in statement that have already been brought up * # of medical terms |  |  |  |
| **Medical Jargon**  Specific jargon terms should be defined, or we need to find a preexisting dictionary | * Dictionary for known jargon terms or phrases |  |  |  |
| **Language complexity** | * Dictionary for filler language (Um, alright, great) * Readability scores * English proficiency score? * Google NLP? * Fragments (How is how is it going?) |  |  |  |
| **Leading questions**  “You don’t have a fever?”  “You’ve been exercising?”  Questions that have an assumption in them | * Negations |  |  |  |
| **Stringed Questions**  “Have you had any fever, nausea, vomiting?” | * “or” * # of medical/key words * Length of question |  |  |  |
| Responds to non-verbal communication | * Video analysis |  |  |  |
| Engages in difficult conversation | * Topic modeling? |  |  |  |
| **Provides reassurance/support** | * “I’m sorry” * sentiment analysis |  |  |  |
| Summarizes possible causes of illness/next steps |  |  |  |  |
| Asks for patient’s questions  “Do you have any questions for me?” | * word “questions” * location in conversation |  |  |  |

**Clinical Skills:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Data to be Extracted*** | ***Features for Model*** | ***Preparation*** | ***Results - Weight Loss*** | ***Results - Abdominal Pain*** |
| Specific items in history | * Use known knowledge on case to determine % uncovered |  |  |  |
| Review of systems questions  Look up list of standard ROS questions | * Dictionary of ROS terms (fever, abdominal pain) |  |  |  |
| Physical exam maneuvers | * Use videos of PE practice to auto-label data and then ML |  |  |  |
| Time/words to diagnosis/diagnostic reasoning | * Could use Isabelle or other diagnostic tool to see how differential changes over the conversation and how the students questions relate to that differential if at all |  |  |  |

**General Features that could be used in models:**

* word counts, syllable counts, words in preceding and next utterance, percent through convo
* Google NLP features (POS, parse label, morphology)
* Counts of medical terms from pre-built dictionary