Question 1: What are some common machine learning algorithms used to predict health insurance costs?

Answer: Regression models, decision trees, ensemble methods like random forest and gradient boosting, support vector machines, and neural networks.

Question 2: What is one advantage of using ensemble methods over individual decision trees?

Answer: Ensemble methods like random forest can comprehend the relations between different features and health insurance costs claimed by the insurance companies, leading to higher accuracy.

Question 3: What is the role of feature selection in machine learning models for predicting health insurance costs?

Answer: Feature selection helps improve the output of machine learning models by deciding which raw input features are essential for the model. It performs dimensionality reduction.

Question 4: What are some common evaluation metrics used to assess machine learning models for predicting health insurance costs?

Answer: Common evaluation metrics include mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), R-squared, and adjusted R-squared.

Question 5: What is the purpose of model validation?

Answer: Model validation is necessary to prevent over-optimism and determine if a model overfits. It assesses the predictive performance on unseen data to build robust models.

Question 6: What is one challenge of using neural networks for predicting health insurance costs?

Answer: Neural networks can have high computational demands and a black-box nature, which poses challenges for interpretability and implementation.

Question 7: What are some examples of feature engineering techniques?

Answer: Examples include deriving new variables from existing ones, modifying variables, pooling historical claims information, and generating interaction variables among predictors.

Fold the paper along the middle to make your flashcards.

Question 8: What are some common validation techniques mentioned?

Answer: Common validation techniques mentioned are cross-validation (especially k-fold cross-validation), bootstrapping, and hold-out validation.

Question 9: What are the main challenges in using machine learning for health insurance cost prediction?

Answer: The main challenges are availability and quality of data - accurate and fast predictions require high-quality and comprehensive datasets. Limited or biased data can affect model development.

Question 10: What is one drawback of linear regression models according to the text?

Answer: Linear regression models do not assume a non-linear form of the relationship within the data, which can make the model less accurate.

Question 11: What are some commonly used machine learning algorithms for health insurance cost prediction?

Answer: Regression models, decision trees, ensemble methods like random forest, support vector machines, and neural networks.

Question 12: What are some challenges and limitations of machine learning models?

Answer: Lack of interpretability, limited explanations of deep learning models, development of wrong or unreliable models, and unlearning of socio-economic and racial biases.

Question 13: What are SHAP and LIME techniques used for?

Answer: SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) are used to attain better interpretable machine learning models and address the challenge of lack of interpretability.

Question 14: What ethical concerns should be considered when using machine learning for healthcare?

Answer: Data privacy, algorithmic fairness, transparency, traceability, and following regulations like HIPAA to protect patient information and retain public trust.

Fold the paper along the middle to make your flashcards.

Question 15: What are some gaps identified in current literature?

Answer: Limited comparisons of ML algorithms, restricted use of multi-modal data, lack of real-world implementation, and need for broader frameworks to address ethics.

Question 16: What is the goal of the current study according to the document?

Answer: To address literature gaps by providing rigorous comparisons of ML methods, pointing out ethical concerns, and assessing predictive models' efficiency for real-world healthcare applications.

Question 17: What types of data can be effectively used by machine learning according to the document?

Answer: EHRs, genomic data, and surrounding environmental data to enhance predictions by considering relationships between genetic susceptibilities, environmental influences and healthcare habits.

Question 18: What evaluation metrics can be used?

Answer: Accuracy, size of data managed, readability/interpretability for different forms of health insurance data and provider settings.

Question 19: What does the document say about existing studies?

Answer: They fail to provide empirical evidence and deploy ML models in working healthcare systems despite theoretical frameworks and experimental validation efforts.

Question 20: What framework can help address aspects of fairness, transparency and traceability according to the document?

Answer: Broader frameworks and regulations are needed to address ethics in ML implementation for healthcare cost prediction.

Fold the paper along the middle to make your flashcards.